DESIGNING DECISIONS

What World of Warcraft and Magic: The Gathering teach about incentives, uncertainty, and behavior in games

Arthur Mota

Designing Decisions

What World of Warcraft and Magic: The Gathering teach about incentives, uncertainty, and behavior in games

Arthur Mota

Designing Decisions: What World of Warcraft and Magic: The Gathering teach about incentives, uncertainty, and behavior in games

© 2026 Arthur Mota. First edition, October 2026.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0), https://creativecommons.org/licenses/by-nc-nd/4.0/.

World of Warcraft is a trademark of Blizzard Entertainment. Magic: The Gathering is a trademark of Wizards of the Coast. This book is an independent analysis and is not affiliated with or endorsed by either company. The cover is an original drawing.

Contents

Games Are Decision Machines **1**
A Rule Is a Situation 2
Two Games That Kept Their Drafts 5
How the Book Is Built 6

I Designing Decisions **9**

1 The Player Does Not Play Your Intentions **11**
1.1 The Rule Paid for the Race 11
1.2 Players Rewrite What a System Is For 12
1.3 The Same Gap at a Card Table 13
1.4 Short Horizons and Loud Losses 14
1.5 Where Incentives Stop Predicting 15

2 More Choices Are Not Better Choices **19**
2.1 Three Talent Systems, One Lesson 19
2.2 A Long Menu Fails Three Ways 20
2.3 A Price Makes a Choice 22
2.4 Options Against Choices at the Card Table 23
2.5 What Cutting a False Choice Costs 24

3 When the Best Strategy Is the Boring Strategy **27**
3.1 The Shortcut Is Worth the Length of the Road 27
3.2 The Team Fixed the Route and Left the Player Alone 28
3.3 Pricing the Dull Route Out Before Anyone Takes It 30
3.4 A Boring Optimum Built From Legal Cards 31
3.5 Dominant, Degenerate, and Shipped Anyway 32

4 Constraints Create the Game **37**
4.1 A Price Leaves the Habit in Reach 38

4.2 A Constraint Takes an Answer and Leaves the Player 39
4.3 Two Factions, Fixed Routes and Bodies That Must Wear Armor 40
4.4 Once Per Turn 41
4.5 The Queen Problem 42

II Uncertainty 45

5 A Solved Game Stops Being a Game 47
5.1 Knowledge Turns a Wall Into a Routine 47
5.2 Players Buy Certainty Wherever It Is Sold 48
5.3 Eight Sources of Doubt 49
5.4 Putting the Unknown Back, or Never Letting It Out 51
5.5 When Solved Is Not Dead 53

6 Random Is Not the Same as Fair 57
6.1 A Correct Generator Builds Its Own Case Against Itself . . . 57
6.2 Memory Is a Dial, and Someone Can Read It 59
6.3 The Price of a Failed Roll 60
6.4 A Deck Is a Bag Sampled Thinly 61
6.5 Some Players Came for the Spike 62

7 Information Is a Game Mechanic 67
7.1 Fog Is What Nobody Decided to Hide 67
7.2 Put the Fact Where the Eyes Already Are 69
7.3 Magic Hides Contents and Publishes Counts 70
7.4 A Secret Needs Edges 71
7.5 The Same Secret Is Play at One Table and Fog at Another . 72

8 Execution Is Different from Knowledge 75
8.1 Knowing the Answer Is Not Performing It 75
8.2 A Shorter Window Turns Two Dials 77
8.3 The Channel Is One of the Performers 78
8.4 Magic Counts Execution as Overhead 79
8.5 When the Noise Is the Game 79

III Designing Encounters 83

9 A Boss Is a Behavioral Machine 85
9.1 Damage Is Not a Question 85

9.2 Six Questions, and What a Missing Answer Makes 86
9.3 The Clock Every Reply Is Counted On 89
9.4 A Zero Anywhere Gives Zero 90
9.5 When the Answer Is Written Before the Pull 91

10 The Gruul Problem: Preventing the Obvious Solution 95
10.1 A Rule a Seating Chart Can Answer 96
10.2 Scatter the Raid, Then Pay for the Scatter 96
10.3 The Slow, the Stone and the Clock 97
10.4 Three Ideas Under the Chain 98
10.5 Poison Took Fourteen Years 101
10.6 When the Chart Is the Fun 102

11 Timer or Health? 105
11.1 A Clock Pays for Gear Twice 105
11.2 The Trigger Decides What the Fight Measures 107
11.3 Clocks That Count What Gear Cannot Buy 109
11.4 A Trigger Has a When as Well as a What 110
11.5 Persistence Protects the Mechanic, Not the Puzzle 110

12 Coordination Changes Everything 113
12.1 The Prototype Failed Because of Its Host 114
12.2 Quality Belongs to the Pair 115
12.3 Remove the Equipment and the Mechanics Break 115
12.4 A Coordination Problem Is Solved by Agreeing 116
12.5 The Principle Will Not Find the Host for You 118

13 When One Player's Mistake Becomes Everyone's Problem 121
13.1 A Mechanic Assumes Cover 122
13.2 Price the Slip and Keep the Stake 122
13.3 Threat Is a Budget the Tank Holds 123
13.4 One Life Total for Two Players 124
13.5 Every Player Added Is Another Chance to Slip 125

14 A Game for Everyone Is Tuned for Someone 129
14.1 One Setting Cannot Serve Two Commitments 130
14.2 Split the Rule and the Problem Moves to the Seam 131
14.3 The Same Split in a Character and a Card Pool 134
14.4 A Menu Works When Each Player Prefers Their Own Item . 135
14.5 When One Rule Serves Both 136

IV Players Against the Designer 139

15 Players Will Find the Tell 141
15.1 Five Tells, and Only One Was Chosen . . . 141
15.2 Looking for the Hole Is Play . . . 142
15.3 What to Call a Find . . . 143
15.4 Cards Have No Creature ID . . . 145
15.5 When the Search Runs on Scripts . . . 146

16 Players Are Better at Finding Problems Than Solutions 149
16.1 Every Complaint Carries Two Messages . . . 150
16.2 What Players Asked For and What They Did . . . 150
16.3 Believe the Report, Audit the Remedy . . . 151
16.4 How Hard to Act on a True Report . . . 154
16.5 When the Players' Fix Is the Right One . . . 155

17 An Unwritten Rule Is Still a Design Decision 159
17.1 Three Places to Settle a Behavior . . . 159
17.2 The Code Where Possible, Custom Everywhere Else . . . 160
17.3 Moving Loot Into the Code . . . 162
17.4 Magic Writes Its Norms Down, Except One . . . 163
17.5 How Far a Norm Holds . . . 164

V Economies, Rewards and Progression 169

18 Loot Is an Economy 171
18.1 A Price Nobody Typed . . . 171
18.2 Players Forecast What Falls . . . 173
18.3 Decide How Rare Before Deciding What . . . 174
18.4 More Rows Than Slots . . . 176
18.5 Where a Fixed Rate Stops Protecting . . . 178

19 Progression Is a Mathematical Promise 181
19.1 Subtraction and Percentages Both Break . . . 182
19.2 Make Survival the Straight Line . . . 183
19.3 A Combat Formula Is a Price List . . . 185
19.4 Mana Is the Denominator . . . 186
19.5 When a Cap Is the Right Answer . . . 187

20 The Economy Remembers 191

20.1 Players Measure from Where They Stand 191
20.2 Too Strong, Too Sure 193
20.3 The Armor That Had to Be Earned Twice 194
20.4 The Cards Players Own 195
20.5 When the Outcry Should Be Waited Out 196

VI Complexity **199**

21 Complexity Has a Budget **201**
21.1 Three Costs on Three Schedules 201
21.2 Rarity Decides Who Pays 203
21.3 Four Buttons and an Auctioneer 204
21.4 Where Hidden Load Comes Back 206

22 Design the System, Not the Feature **209**
22.1 Every Piece Waits on Another 210
22.2 An Owner, a Tool, a Multiplier and a Test 211
22.3 The Skeleton Before the Cards 212
22.4 The Order of Construction Is a Design Decision 213
22.5 When the Play Overrules the Plan 215

VII Iteration **219**

23 Most Good Mechanics Begin as Bad Mechanics **221**
23.1 The Verdict and the List 221
23.2 Reasons That Traveled 223
23.3 Energy, Chroma and the Rules for Failing 224
23.4 Not Every Bad Mechanic Is Waiting 225

24 Technical Constraints Are Game Mechanics Too **229**
24.1 Where the Data Sits Decides What Can Be Played 229
24.2 A Door Is as High as the Camera Allows 231
24.3 What a Limit Costs 232
24.4 The Card With No Back 232
24.5 When the Limit Is a Guess, and When It Has Moved 233

25 Designing a Game That Will Change **237**
25.1 A Launch Is a Point on a Line 237
25.2 Good Releases Add Up to a System Nobody Chose 238

25.3 Rotation, Bans and a League a Year Ahead 240
25.4 What Players Have Built on It 241
25.5 Insurance Has a Premium 242

The Designer's Real Job **247**
The Same Order of Events 247
A Fix Moves the Cost 248
The Rule Is the Interface 249

Video Sources **253**

Games Are Decision Machines

A healer in *World of Warcraft* has one job, and the job is never finished. In a raid, where a large group fights a single enormous enemy, a few players called tanks stand in front of the boss and take its attacks, most of the group deals damage, and the healers watch a column of health bars and refill whichever one is draining. A healer who stops casting is a healer about to watch a tank die.

That makes the question behind Loatheb a strange one. Loatheb is a boss in the Naxxramas raid, and his designer, who publishes commentary on his own encounters under the name Candlesan, says the fight started from a single question: could the team make an encounter in which the healers would be called on to deal damage?[1] The people who had come to heal would spend long stretches doing the one thing their role never does. A fight that simply needed fewer healers would have missed the point.

The first answer takes one line. Turn healing off. The team dropped this naive design quickly, for a reason that had nothing to do with combat. Disabling healing would not make healers deal damage. It would get them benched.[2] A raid leader filling a roster for a fight with no healing in it leaves the healers outside and invites players built to hit things. The rule was aimed at what a healer does during the fight. It landed on a different decision, made by a different person, before the fight began.

So the team tried a price instead of a ban. In the first real attempt, heals placed a stacking penalty on the tank, and consecutive heals were worth less. Healers were supposed to space their heals and spend the gaps dealing damage.[3] Testing produced the reverse. When the tank's health got low, everybody panicked and healed harder, trying to push through the penalty by volume. "We got more healing, not less."[4]

The second attempt moved the penalty from the tank to the healer. Each healer carried a personal counter, and every heal made that healer's next

one weaker: the first at full strength, the second at 80 percent, the third at 60 percent.[5] On paper this repairs the first version. The cost belongs to one person, that person can see it, and nobody can hide in the crowd. It was miserable to play. Sit at three stacks, so that your next heal lands at 40 percent, less than half of your first. The tank needs a heal. "Doing nothing feels like a mistake." Casting feels like digging a deeper hole.[6] The rule offered a choice and made both branches of it feel like an error.

The version that shipped is cruder than either, and it worked. One heal per healer per minute.[7] It has no slope, no stacks and nothing to estimate. Raid leaders wanted healers in the group, because each one carried a heal the raid could not do without. And a healer who had just spent that heal had nothing useful left to do except attack.[8] The fight that began as a question about healers dealing damage got its answer from a rule that never mentions damage.

Four rules were aimed at one behavior. The ban changed who got invited. The shared penalty produced more of the thing it taxed. The personal penalty produced a decision nobody wanted to make. The cap, blunt enough to be read at a glance, left each healer with most of a minute and one sensible way to spend it (Figure 1).

A Rule Is a Situation

The four versions do not differ in size, and they do not differ in purpose. What moved was the situation each rule put a person in. That is the thesis of this book: designers do not design rules so much as environments in which people decide, and the decision, not the rule, determines what happens next. A game is a ***decision machine***. The designer builds the situation, the player supplies the choice, and the game is whatever comes out when a great many players supply theirs at once.

A single rule changes that situation along six lines, and Loatheb shows all six.

It changes ***incentives***, meaning what the rule pays for as opposed to what its author hoped for. A weakened heal is a partial loss and a dead tank usually ends the attempt, so as long as more healing could still save the tank, more healing was what the shared penalty rewarded. The count was shared, too, so restraint paid only if every healer practiced it, and a healer who waited while the others cast gave up a heal and gained nothing. In mechanism design a rule is incentive compatible when following the strategy it prescribes is the best thing a participant can do under it. The shared

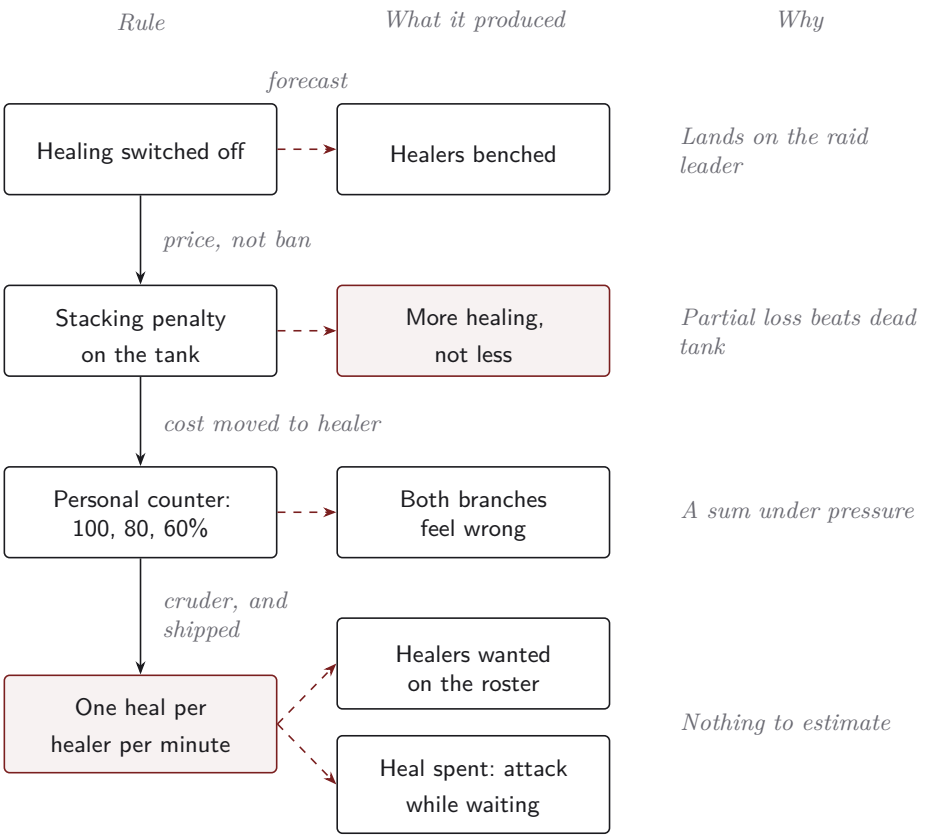


Figure 1. Loatheb's four rules, in the order they were tried. The goal never moved; each rule changed the situation a person faced. The dashed arrows mark what followed from each rule, the first of them foreseen and never tested.

penalty asked for restraint and paid for panic. A harsher penalty would not have fixed it. While a weakened heal still does anything for a dying tank, the ranking holds, and at the setting where it does nothing the price has become the ban the team had already thrown out.

It changes ***information***, meaning what the player can read at the instant of choosing. The health bar is large, it moves, and what it threatens is final. The stack count is a small number whose meaning has to be worked out, and the personal counter asked for that sum at the worst possible moment. The cap asks a question with two answers. Either the heal is available or it is not.

It changes ***constraints***, the set of actions that exist at all. A penalty leaves every action available and alters its cost. A constraint removes actions. The naive design and the shipped design are both constraints, and the difference between them is one heal. Removing every heal removed the reason to bring a healer. Removing all but one made that one precious.

It changes ***opportunity costs***, what the player gives up by acting. Under both penalties, a healer who turned to attack gave up attention to a tank who might need help at any second. Under the cap, a healer with the heal spent gives up nothing by attacking, because the alternative is gone.

It changes ***expectations***. Healers arrive at Loatheb with a habit, and the sense that doing nothing is a mistake comes from that habit, not about the numbers. A rule that contradicts what players believe their job to be has to overcome the belief as well as the payoff.

And it changes ***social behavior***. The first decision any of these rules touched belonged to the raid leader. Benching is what a rule does when its effects arrive at the roster before they arrive at the fight.

No rule moves one of these and leaves the rest alone. The cap is a constraint, a change in opportunity cost, a simplification of what the healer must read, and an argument for keeping a seat filled. Its consequences did not stop at the healers either.

Second-Order Effect: The spores

With the cap in place, the next worry was that less healing meant tanks might not get the rage they needed to generate enough threat.[9] The answer was a new object in the room: spores that, when killed, give players a period of dealing damage that generates no threat at all.[10] That answer had its own cost. The fight would go on requiring consumable potions even after a raid had outgrown it, which is unusual,

> and its designer says plainly that he does not love that outcome.[11] A fix for healers became a problem for tanks, and the fix for tanks became a standing bill for everyone.

The two failed penalties share a mechanism. A graded penalty hands the player an estimate to perform, and people under time pressure answer the most visible fact in front of them long before they finish the estimate. A hard stop can be seen. The same lesson turned up at the scale of the whole game when cooldowns were briefly taken out of *World of Warcraft*'s combat, the meeting Chapter 4 recounts.

The thesis has a limit, and it should be stated once. Calling a game a decision machine is not an instruction to add decisions. The claim is narrower: whatever a designer builds, a decision is what the player receives, so the decision is the thing to inspect.

Two Games That Kept Their Drafts

Testing a claim like that takes games whose designers wrote down what they tried before the version players remember. About two thirds of the cases here come from *World of Warcraft* and the rest from *Magic: The Gathering*, and no other game appears except where one of these two inherited a mechanic from it.

World of Warcraft earns its share because its record of first attempts is unusually deep. John Staats, a level designer, kept a diary through the years before release, and designers kept explaining their reversals in public long after. That record holds quests that began as navigation aids and became the game, and heals that drifted into being castable on the run until the team put a price on moving back (both in Chapter 1).

Magic is the second laboratory because its physics are the opposite. It is played in turns, across a table, with no reflexes involved, and most of what a player does not know sits in a shuffled deck or an opponent's hand. A principle that holds in a real-time fight among a large group and also in a card game for two is not an accident of genre. The two games share a preoccupation all the same. Mark Rosewater, who has written about *Magic*'s design for years, separates options from choices, and an added option is not an added decision (Chapter 2). He also supplies the counterweight: sometimes the right move is to stop designing the decision and adopt the one players have already made (Chapter 1).

Neither studio is an oracle. Most of the cases are things that did not work the first time, and some of the repairs are open to argument. Players who find the efficient response to a rule are not misbehaving. They are the experiment, and in case after case they show the designer what the rule said.

None of this requires having played either game. Where a case turns on a mechanic, the mechanic is explained at the point it matters. A reader who has spent years in one of these games will recognize the fights and the cards. A reader from economics or decision theory will recognize something else: a long series of design experiments in which someone changed one rule, a large population responded, and the person who changed it wrote down what had been expected.

How the Book Is Built

Every chapter opens on one concrete case, follows what it shows into a second case that sharpens it, states a principle where the cases earn it, and marks where that principle stops working. Each ends with a short set of questions, the Designer's Test, which has no answer key. The sources sit in the notes at the end of each chapter.

The seven parts follow the questions in the order a designer meets them. Part I stays with decisions: why players do something other than what the designer pictured, when a choice adds anything to play, what to do when the best strategy is the dull one, and why removing options so often improves a game. Part II turns to uncertainty, including a case in which the arithmetic is correct and the game feels broken anyway. Part III is about encounters, and ends by asking whose game a shared encounter is tuned for. Part IV follows players who beat fights with tricks their designers never anticipated, and the rules players keep for each other where the designer wrote none. Part V treats loot and progression as economies with memories. Part VI asks how much complexity a game can afford. Part VII is about iteration, technical limits, and designing a game that will keep changing after it ships. The conclusion returns to Loatheb.

Underneath the parts is a single chain, and each chapter walks some length of it. A designer writes rules. The rules create incentives, whether or not anyone intended them. Players facing those incentives make decisions. Decisions repeated by many players become behavior. Behavior, in a system with many moving parts, becomes outcomes that nobody wrote down,

including a fight whose healers stop healing, and a raid leader who invites them anyway.

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V6 at 0:02.
2. Candlesan, commentary, V6 at 0:12.
3. Candlesan, commentary, V6 at 0:24.
4. Candlesan, commentary, V6 at 0:39.
5. Candlesan, commentary, V6 at 0:48.
6. Candlesan, commentary, V6 at 1:06.
7. Candlesan, commentary, V6 at 1:25.
8. Candlesan, commentary, V6 at 1:31.
9. Threat is the game's running account of which player an enemy most wants to attack. A tank's job is to stay at the top of it.
10. Candlesan, commentary, V6 at 1:43.
11. Candlesan, commentary, V6 at 2:00.

Part I

Designing Decisions

Chapter 1

The Player Does Not Play Your Intentions

In September 2002 the team building *World of Warcraft* opened an internal alpha to the rest of Blizzard, and the company played one starting zone for the better part of a week. Allen Adham's email to the company about the test was a happy one. It counted 61 users at the Friday peak, which he called an all-time high, and it listed among the few blemishes a bit of kill stealing and a whole lot of ninja looting.[1]

The looting had a cause, and John Staats, whose diary of the game's development is the record here, gives it in one line. "Anyone could loot any corpse, which meant melee players got everything." Rogues came out of stealth to loot monsters that other people had killed.[2] The melee characters won because they were standing next to the monster when it fell.

Then the testers changed what they were trying to do. Playing that way was frustrating, Staats writes, so everyone's aim became to raise a character's level, get out of the congested starting area and play solo.[3]

Nobody had designed a race. No rule said that the first player to click owned the kill. The absence of a rule said it. The testers answered in two steps, and both were sensible: grab what you can while you are in the crowd, and leave the crowd as fast as you can. For that week, a game being built around other people paid best to anyone who avoided them.

1.1 The Rule Paid for the Race

Strip the zone away and the problem has three lines. The designer wants behavior A. The rules reward behavior B. Players choose B.

The word ninja puts the blame on the looters. But a tester who declined to loot under that rule simply went without, and nothing in the game thanked anyone for the courtesy. The behavior was the rule, played well. Chapter 17 returns to that habit of blaming the player for the rule.

The working idea is the ***revealed incentive***: whatever a rule pays for, read off what players do under it and not off the design document. The name borrows an economist's habit of inferring what people want from what they choose. Open looting had no stated purpose at all, and its revealed incentive was to stand close, click fast and then go somewhere empty.

Players respond to the game you built, not the game you meant to build.

A design document is a hypothesis about what a rule pays. Play is the measurement. That makes the principle a diagnostic before it is advice: when a mechanic produces behavior nobody intended, the first question is what the rule is paying for, and what players have misunderstood comes second.

> **Second-Order Effect: The test that tested the wrong game**
>
> Open looting did more than hand out loot unfairly. Staats calls it a perfect example of one oversight muddying the entire gameplay experience, and adds that most of the feedback about playing with other people was a bit spoiled as a result.[4] The alpha was the team's instrument for learning whether grouping was fun. The loot rule gave testers a reason to treat nearby players as competitors and then to get away from them, so what the instrument measured was the loot rule. A rule with the wrong incentive distorts the evidence a designer would use to judge every rule near it.

1.2 Players Rewrite What a System Is For

Quests were not built to be the way a character levels. They began as navigation breadcrumbs. Designers assumed players would clear a zone's quests to learn the place and then grind monsters there until they were ready to move on. Testers with an empty quest log left early to refill it, into zones too hard for them, and the only way to hold them was to write far more quests than planned. The solo content that resulted was never planned, budgeted or prioritized. "By trying to solve a navigation problem, we'd inadvertently engaged a larger audience."[5] Years later the arrangement had settled: designers knew how players would behave in zones, and exterior level designers worked around it.[6] The players did not come around to the plan. The plan was rebuilt around what the players did.

The quests then taught the team two more prices. Designers believed that players who reached level 40 and received a mount would enjoy quests that sent them across a continent. Travel time turned out to be downtime. Jeff Kaplan, of the quest team, took the idea back in front of his colleagues, and the designers settled on short-range quests clustered around hubs, keeping only a few long journeys.[7] The designer had pictured the journey. The player was watching a clock, and on the clock a ride across a continent was a long stretch in which nothing was earned.

Dungeons taught the second. The team did not know how long clearing a dungeon would take, or whether players cared about clearing one at all. Uldaman, built by Dana Jan, was not linear, so its quests required several visits. Only when it had been played by people who were not developers did the team learn that players disliked leaving a dungeon unfinished and wanted the closure of a full clear.[8] No reward table contains that preference. Players count an unfinished place as a cost, and a layout that looks like generous replay value on a floor plan hands them that cost at the end of every visit.

The third case had its numbers in plain sight. Boss fights make players move, and movement is supposed to cost a healer casting time. A Blizzard design post on healing said that healers had come to cast many of their heals while moving, that the team wanted movement to cost them something again, and that it wanted players fighting other players to have more ways to limit healing than silences and crowd control. Several heals, Prayer of Mending and Wild Growth among them, went from instant to a cast time of 1.5 seconds, a second and a half of standing still where before there had been none.[9] Offered a heal that costs nothing to cast on the run next to one that roots them in place, healers who had to move used the first, and the encounter designer's price for movement went unpaid.

A long road, an unfinished dungeon and a free heal have little else in common. In each, the designer priced the rule in one currency (grandeur, replay, convenience) and the player paid or collected in another (minutes, closure, throughput).

1.3 The Same Gap at a Card Table

Magic: The Gathering has no health bar draining in real time. Its players sit across a table and take turns, and nothing rewards a fast hand. Mark Rosewater's columns on its design record the same gap anyway, three times, with three different causes.

The first is a model. Suspend, a mechanic Rosewater dates to 2006, let a player pay less for a creature in exchange for waiting several turns for it to arrive, and in *Magic* a creature that has just arrived cannot attack unless it has an ability called haste. Players attacked with their suspended creatures on arrival regardless. The designers first tried to repair the players. They reworded the cards, added visual cues, and considered having the creature enter after combat. None of it held, and the fix was to give every suspended creature haste. Rosewater's lesson is to change the game to match the players and not the reverse.[10] The rewording and the cues restated the same rule more loudly, on the theory that players lacked information. They had a model instead (the wait is over, so the creature is ready), and no wording dislodged it.

The second is pure incentive. For *Unhinged*, a set in *Magic*'s Un- series, Rosewater built a mechanic called gotcha out of an older card that penalized saying certain words: when an opponent slipped and broke a verbal condition, a player could call gotcha and get a card back. Internal playtests were fun. A colleague warned that the mechanic might simply tell players to stop talking, and the warning was right. The designers had been playing along with the joke, while players who put winning first played to the incentive, which made for a quieter and worse game. Players trust the designer to have made something fun, and so they do whatever the game rewards.[11] The mechanic was not repaired, and Rosewater ranks the set as the worst of the four in its series.[12]

The third is fear. Ulamog's Crusher was a common creature in a set whose giant monsters were meant to feel alien and hungry, and it was an 8/8, with four times the power and four times the toughness of the 2/2 flier Rosewater uses as an example elsewhere.[13] In playtests it rarely attacked, because less experienced players are more afraid of a poor choice that loses them a creature. The designers added a line of text that forces the Crusher to attack.[14] The same bias shows up in a smaller card. The beginner overvalues life, and so plays Venerable Monk, a small life-gaining creature, at the first chance, while the expert holds it.[15] In both cases the card's mathematics said one thing and the person holding it was looking at something else.

1.4 Short Horizons and Loud Losses

Two ideas from the study of decisions account for the whole chapter.

The first is ***local optimization***. Every player in these stories solved a real problem well, and the problem was smaller than the designer's. The

alpha tester solved the crowd at the starting area. The rider solved the next half hour. The Crusher's owner solved how not to lose an 8/8 this turn. Each answer is reasonable inside its own horizon and wrong only against a horizon the player was never standing at: the purpose of the test, the shape of the zone, the theme of the set. No amount of effort would have put them there. A tester has no reason to protect the quality of the feedback an alpha was built to collect. Decision theorists call the general condition bounded rationality, after Herbert Simon's behavioral model of rational choice.[16] People choose with limited time, attention and information, settling for what is good enough instead of solving the designer's whole problem, and what they have is always less than the designer had while writing the rule.

The second is ***salience***, with its companion, asymmetric loss. When time is short, the most visible and most catastrophic state sets the decision. A creature that might be lost outranks the damage it might deal, as a dying tank outranked a stack count in the introduction's first Loatheb penalty. Loss weighs more than gain of the same size, and a loss that can be seen weighs more than one that has to be computed. The designer's payoff is counted over the fight or the set, and the player's is counted in the moment.

Together they predict behavior without any knowledge of what the designer intended.

Design equation (a conceptual framework, not a law)

observed behavior = best response to the payoff the player can see, at the horizon the player occupies
intended behavior = best response to the payoff the designer computed, at the horizon of the whole design

A mechanic is in trouble wherever the two lines give different answers. The loot rule gave different answers from the first corpse, and the instant heals did so every time a boss made the raid move.

1.5 Where Incentives Stop Predicting

If players simply followed incentives, two rules with the same incentive would get the same response. Rest experience, which Chapter 20 tells in full, broke that: the same ratio between a fresh and a tired player was hated as a penalty and welcomed as a bonus. People judge an outcome against a starting point, the reference dependence of Kahneman and Tversky's prospect theory.[17]

The revealed incentive predicts what players will do with a rule, not how they will feel about it, so a rule has to be checked twice: once for what it pays and once for the point from which players will measure it.

Suspend marks a second limit. With suspend the payoff was never the trouble. Players held a belief about what a long wait had earned them, and the designers wrote that belief into the rules. Sometimes the player's model is the fixed point and the design is what moves.

The third limit is a cost. Forcing the Crusher to attack removed a decision to protect a theme, and a player who is never allowed to hold the creature back never learns when holding back is right. The principle tells a designer where intention and incentive have come apart. It does not say which of the two should yield.

The fourth is the population. A system is pushed to its extremes only when every player acts as a rational actor playing to win, and much of a *Magic* table plays for the evening and feels little pull toward the extreme.[18] The pressure a rule produces depends on the share of players who optimize. What a designer does with a population that mixed is the question of Chapter 14, and what the players who do not optimize hold in place is the question of Chapter 17.

The Designer's Test

1. If you were maximizing your own payoff, and not playing the fantasy the designer imagined, what would you do with this rule?
2. What is the worst outcome of obeying the rule, and what is the worst outcome of ignoring it? Which is larger?
3. Which state will the player be staring at when the decision arrives, and is it the one your rule changes?
4. Could this rule change what players use the whole system for, or what your next playtest is able to measure?
5. Is the player's model of the rule wrong, or is your rule fighting a model you cannot change?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "September 2002: Internal Alpha 1.0".

2. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
3. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
4. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
5. Staats, *The WoW Diary*, section “Quests”.
6. John Staats. AMA Recap: John Staats, Blizzard’s First 3D Level Designer, on Early WoW Development. Wowhead, recap by Perculia, 2018, last answer.
7. Staats, *The WoW Diary*, section “Quests”.
8. Staats, *The WoW Diary*, section “March 2003: Internal Alpha 3.0”.
9. Blizzard Entertainment. Dev Watercooler: Building Healthy Gameplay. `worldofwarcraft.blizzard.com`, n.d., section “Instant-Cast Heals”.
10. Mark Rosewater. Twenty Years, Twenty Lessons, Part 1. Making Magic, `magic.wizards.com`, 2016, section “Lesson 1”.
11. Mark Rosewater. Lessons Learned, Part 1. Making Magic, `magic.wizards.com`, 2023, section “Unhinged”.
12. Rosewater, “Lessons Learned, Part 1”, section “Unhinged”.
13. The two numbers are a creature’s power (the damage it deals) and its toughness (the damage it takes to destroy). Common is the game’s most frequent rarity.
14. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section “Lesson 14”.
15. Mark Rosewater. Lenticular Design. Making Magic, `magic.wizards.com`, 2014, section “Rule 5”.
16. Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.
17. Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, 1979.
18. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 10 at 12:56 and 13:20.

Chapter 2

More Choices Are Not Better Choices

The first talent system in *World of Warcraft* was a set of tiny red buttons. John Staats, whose diary of the game's development records all three versions of the system, dates its history from November 2002 to June 2004 and gives the reason it existed. A player who gained a level did not feel immediately better. Nothing new arrived until the character had gone back to town and talked to a trainer, and that left a delay between earning the level and feeling improved. Allen Adham thought it would be a fun perk to let players pick a stat to raise, so the buttons went in beside the core stats, one press for each new level.[1]

The method felt good for a while. Then the long-term problems showed. "All the melee classes picked the same stats, no one wanted spirit," and the customization gave players no new ways of playing.[2] Staats calls the result a largely transparent reward, especially at higher levels: a reward the player looked straight through, one more point of strength on a character who already had a great deal of it.

Each press was a decision in the plain sense: several stats on offer, one picked, the character sheet changed. The designers had attached a choice to every level. They had not attached anything a player would notice having chosen.

2.1 Three Talent Systems, One Lesson

The second version offered different kinds of things to choose. By the middle of 2003 a newly leveled character no longer put a point into a stat. Characters learned general abilities and sharpened their combat performance, on the thinking that something like increased damage against a type of monster would be more rewarding than a point of strength.[3] The list ran from weapons and schools of magic to the nearly useless bonus against dragons.

It felt better, and it failed the same way. Mages still poured their points into intelligence and warriors into strength, and a player who chose something odd found that the experience of playing was exactly the same.[4] In July 2003 the talent design was declared finished. Staats adds a spoiler in parentheses: it was not.[5]

The third version was not in the plan. A talent skill tree became possible because the team had hired another senior designer, Tom Chilton, late in development. Each of the nine classes would get its own tree and many new abilities, "about a hundred times more work than the original talent system." Chilton needed a month of very long hours to finish the first two classes, with three months left before the final candidate build.[6] Two public betas later, four of the nine classes had talents.[7]

What the hundredfold cost bought fits in a clause: the skill tree made a player's experience of the game different from that of others in the same class.[8] The first two versions had given every player a choice at every level. Only the third gave two warriors a reason to play differently. Version two is the move every designer reaches for first: if the menu is too thin, write a longer one. A longer list looks like more freedom, and every item on it can be defended on its own. Version two had more kinds of option than version one and the same sameness.

2.2 A Long Menu Fails Three Ways

The game kept lengthening the menu after release. A talent point for every level worked for a game with 60 levels and less well for one with 90, half again as many,[9] and by the *Cataclysm* expansion the stated goal for talents was to remove filler and strengthen specialization.[10] The talent record holds an example of each way a long menu fails.

The first is the ***dominated option***: one that some other option beats in every situation that matters to the player choosing. Spirit on a melee character is the type. The later trees had the same fault at a finer grain: Greg Street, the lead systems designer, judged small incremental talents with a mathematically right answer not a great model.[11] If one node gives slightly more damage than its neighbor in every fight, the tree has asked an arithmetic question, and arithmetic questions have one answer.

The second is filler: nodes that everyone takes. A talent every member of a specialization picks is a toll on the road to the talents below it. It adds a click and a line of text, and it differentiates nobody.

The third is subtler, because the options really are different. Picking the talent you want from among talents that do not interest you is an easy multiple-choice test, and a Retribution paladin offered a damage talent will take it.[12] A healing talent offered to a player who never heals is the same thing. The outcomes differ, and the player's role has answered the question before the player reads it.

When a choice has an answer, players find it once and pass it around as the cookie-cutter build, a tree filled the same way by everyone of one specialization, usually from a guide. Nonstandard builds are usually weaker, and the root is the number of combinations: some interactions beat the rest, and a tree of 40 to 50 talents cannot be fun, interesting and balanced all at once. In many *Cataclysm* specializations, one to four points out of the whole tree were left to place freely.[13] Knowing the right build does not mark a skilled player either, since the build is copied from the web.[14] Copying is not the fault. A solved tree hands players a problem with a single solution, and they use it, as Chapter 1 would predict.

A choice adds gameplay only when its outcomes differ, the player cannot simply look up the answer, the trade-off is real, and the best option depends on the situation.

Each clause removes a false menu: the transparent reward, the cookie-cutter build, the option nobody refuses (or nobody takes), the answer solved once for good. Version three adds what the list leaves implicit: a choice is real when it changes what the player does, not only what a number says.

The conditions hold together as a product of three factors, a framework for asking the question and not a measurement.

Design equation (a conceptual framework, not a law)

$$\text{decision quality} \approx \text{trade-off} \times \text{uncertainty} \times \text{agency}$$

Trade-off is what the player gives up by choosing, the third clause. Uncertainty is the part of the answer that depends on the fight, the opponent or the draw, which covers the second clause and the fourth. Agency is the difference the choice makes to what the player then does, the first clause together with the lesson of version three. A zero anywhere gives zero overall. A ***false choice*** is a menu with one of the three factors missing.

2.3 A Price Makes a Choice

The instant heals of Chapter 1 were a trade-off at zero. Move or heal is a question only while each excludes the other, and the designers restored it with a price, not an option.

The same repair appears in fights between players. Spellsteal lets a Mage take a beneficial effect off an enemy and keep it,[15] and for *Mists of Pandaria* its cost was raised significantly.[16] A nearly free steal is always right to press, which involves no judgment. At a real price the Mage asks which effect is worth the mana, and the opponent reasons about which are safe. A price turned a reflex into a ranking.

The same post reports a failure of the opposite kind. Several Mage control abilities had been moved into talents, so that a player who could take only one would give the others up. It did not produce enough exclusivity, and Mage control and burst were cut further by direct changes to individual spells.[17] The talent screen said choose, but the abilities left behind substituted for one another. Opportunity cost, the introduction's name for what a player gives up by acting, is measured in play, and the layout of the menu does not measure it.

A character's buttons are a menu too. The team has no magic number of rotational abilities, the ones a player cycles through in an ordinary fight, and finds about four to be the sweet spot.[18] The original classes leaned on one ability while many others sat unused,[19] and action bars filled with situational abilities, consumables, mounts, trinkets and macros.[20] A later pruning merged or removed cooldowns that were typically used together.[21] Two buttons always pressed together are one decision counted twice, and a neglected ability is a dominated talent: on the screen and in no one's play. Why the bars stay full, and who pays for it, is the subject of Chapter 21.

A choice can also fail by being too strong. The pandaren got racial bonuses meant to feel unique without forcing the pick, a cooking bonus and reduced falling damage,[22] while one existing racial was very powerful and the high outliers were brought down.[23] If one race is plainly stronger, the choice of how a character looks is made by a spreadsheet.

Second-Order Effect: Infinite dungeons

Option count has a cost that no single player's menu shows. In July 2001 the procedurally assembled dungeons of *Anarchy Online* were immediately popular with the game designers, and it was decided that

> an infinite number of generic dungeons would beat a smaller number of unique ones, sacrificing flavor for quantity.[24] By October the designers had abandoned the approach and agreed on a total of thirty dungeons. One deciding factor was that too many dungeons "might divide the player base in what was supposed to be a social game."[25] A dungeon needs a group, and every destination thins the supply of people heading to the others. More options for each player meant fewer players in each option.

2.4 Options Against Choices at the Card Table

Mark Rosewater's 2009 essay on decisions in *Magic: The Gathering* opens with two imaginary games, one with too many rules to play and one won at once, and aims for a middle that depends on who is playing.[26] It draws a line between options and choices: an option lets a player decide what to do, a choice makes the player decide how, and choices produce the interesting states of a game. His example is an invented creature whose owner can pay to give it first strike. Give it a second payable ability, trample, and it becomes more useful without becoming more interesting, because its owner takes whichever fits the moment and nothing is traded. Rebuild it as a 4/4 that pays for first strike by shrinking permanently, and both players have something to weigh.[27][28] Even that version's effect is somewhat limited, he grants: a real trade-off made the card better, not every activation a dilemma.

His second example was printed. Expunge, from *Urza's Saga*, is a strong removal spell, one that destroys an opposing creature, and it was given cycling, which lets a player pay to discard the card and draw another. Cycling it is correct in three narrow situations, such as being stuck at two mana, and almost never otherwise. Rosewater calls it a design failure, and his standard is that the choice a mechanic offers should be neither almost always taken nor almost never.[29] It is spirit in a card game: dominated everywhere except three corners.

Kicker lets a player pay extra when casting a spell to get more from it, and over time it came to cover nearly anything a card could offer for more mana, until players greeted each new extra-cost mechanic as kicker under another name. Rosewater calls the mistake making a mechanic too flexible. The repair restricted kicker to two outcomes, a bigger effect or a second effect that works with the first, and gave everything else its own keyword.[30] A dozen decisions had been filed under one name, and the name stopped saying which was on offer.

Decisions made before the game begins follow the same rule. Limited play, where decks are built from cards drafted on the spot, is meant to encourage discovery, with players stretching for weaker cards late in a draft.[31] Scarce strong cards force the pick. A draft of all good cards would have many options and few choices.

The principle's second and fourth clauses have a name among *Magic* commentators: choice variance, one piece with different best uses in different states of the game. A card with kicker is one spell early and another late, so its holder has to judge which the moment wants. Its mark is a rule of thumb that is usually right and wrong often enough that the skill under test is knowing when to break it.[32] A cookie-cutter build has a lookup and no rule of thumb. A choice with a rule of thumb and no lookup is live.

Decision density is the number of real decisions a player faces in a stretch of play, as against the options on display. Four rotational buttons and Rosewater's two impossible games both say it has a best range: too few and the game plays itself, too many and the player follows a script. Count what survives once the dominated, the mandatory and the already answered are struck out. In the *Cataclysm* trees, that was one to four points.

2.5 What Cutting a False Choice Costs

If false choices add nothing, cutting them should cost nothing. *Cataclysm* tested that. The team pruned the filler and moved the choice of specialization to level 10, where it became a formal decision. The level 10 choice was close to universally acclaimed. The rest read as content lost for little gain, did not fully solve its problems, and *Mists of Pandaria* redesigned talents again, toward more impactful talents and more freedom to change them.[33] Players greeted that redesign by complaining that they would have fewer choices, the exchange Chapter 16 follows.

The players were wrong about the count and right about something else. A point spent in a filler talent is still a small event at a level, and supplying that event was version one's entire purpose. People measure a change from where they stood (Chapter 20), so a false choice has a perceived value, and removing it is a loss even when the remaining decisions improve. The one part of the overhaul players praised added a decision. The parts they resented only subtracted.

Two further limits hold. Solved choices do not go away: some throughput talents in the new system were expected to keep a clear answer for a given encounter, and the team accepted it.[34] And the clause about context cuts

both ways. Talents built around slowing or controlling an enemy meant nothing to players who never fought other players, since raid bosses ignore such effects, and the sameness of the outdoor encounters was a problem with the game.[35] An answer that depends on the situation is a live choice only for a player who meets more than one situation. For everyone else it is a dominated option that looks like a real one.

The Designer's Test

1. Is there an option a fully informed player never picks?
2. What does choosing A cost the player in B, and is that cost paid in play or only shown on the menu?
3. Does the best answer change from one fight, match or opponent to the next, for the players who will actually face the choice?
4. Does the choice change what the player does, or only a number?
5. If you removed this option, would anyone's play change, and would anyone feel robbed?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "June 2004: Public Beta 2.0".
2. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
3. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
4. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
5. Staats, *The WoW Diary*, section "July 2003: Unexpected Giants".
6. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
7. Staats, *The WoW Diary*, section "August 2004: Public Beta 4.0".
8. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
9. Greg Street. Seeing the Forest for the Talent Trees. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section 4.
10. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What didn't work out as planned?".
11. Street, "Cataclysm Post Mortem", section "What didn't work out as planned?".
12. Street, "Talent Trees", section 1.
13. Street, "Talent Trees", section 2.
14. Street, "Talent Trees", section 5.
15. A beneficial effect, or buff, is a temporary spell on a character that makes it stronger. PvP, player versus player, is the part of the game in which players fight each other and not monsters.

16. Greg Street. Dev Watercooler: Mists of Pandaria PvP. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "Mages".
17. Street, "Mists of Pandaria PvP", section "Mages".
18. Greg Street. Dev Watercooler: Number of Abilities. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What's the Magic Number?".
19. Street, "Number of Abilities", section "What's the Magic Number?".
20. Street, "Number of Abilities", section "The Human Factor".
21. Blizzard Entertainment. Dev Watercooler: Pruning the Garden of War. `worldofwarcraft.blizzard.com`, n.d., section "Ability Pruning".
22. Blizzard Entertainment. Blizzard Insider 42: Meet the Pandaren Monk (interview with Greg Street). `worldofwarcraft.blizzard.com`, n.d., first question.
23. Blizzard, "Pruning the Garden of War", section "Racial Traits".
24. Staats, *The WoW Diary*, section "July 2001: Nine Months Down the Tubes".
25. Staats, *The WoW Diary*, section "October 2001: Learning from the Good and Bad".
26. Mark Rosewater. Decisions, Decisions, Part I. Making Magic, `magic.wizards.com`, 2009, introduction.
27. Rosewater, "Decisions, Decisions", section "Level I, The Card Level".
28. First strike lets a creature deal its combat damage before the opposing creature does. Trample lets excess damage carry through to the defending player. A 4/4 has four power and four toughness.
29. Rosewater, "Decisions, Decisions", section "Level II, The Mechanics Level".
30. Mark Rosewater. Here's the Kicker. Making Magic, `magic.wizards.com`, 2007, section "Kicked When You're Down".
31. Mark Rosewater. Nuts & Bolts 2: Design Skeleton. Making Magic, `magic.wizards.com`, 2010, section "#1".
32. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 3 at 5:37; episode 15, part B, at 0:14 and 3:52.
33. Street, "Cataclysm Post Mortem", section "What didn't work out as planned?".
34. Street, "Talent Trees", section 3.
35. Street, "Talent Trees", section 6.

Chapter 3

When the Best Strategy Is the Boring Strategy

A monster in *World of Warcraft* does not see the world a player sees. Besides the geometry drawn on the screen and the invisible geometry that keeps characters from falling through floors, the engine kept a third kind. Pathing geometry tells monsters where they can run, and the game generated it by heuristic as a patchwork of floor tiles. The result was a rough copy of the place: in one printed example a staircase has been flattened to a plane and a door is missing altogether.[1]

A rough copy has gaps. Somewhere in a cave or a keep there is a ledge where a character can stand and the monster's tiles do not reach. From there the player attacks and the monster cannot answer. The team called such a place an immunity spot.[2]

Closing the gaps took years. The programmer Scott Hartin rewrote the pathing code more than a dozen times.[3] In April 2002 the team tried steering monsters along splines, curves in place of straight lines, and dropped the idea because nothing controlled how sharp or wide a curve should be.[4] In October 2002 the code looked as if it might have to be rewritten from scratch again.[5] Stable pathing inside buildings and caves arrived in June 2003, after a year of on-again, off-again work, and only then could designers place monsters in dungeons and script fights there.[6]

No tester had yet been caught in an immunity spot. The team spent the years because it expected someone would, and a player who could attack without taking damage could repeat the trick for free loot and experience, damage the economy, and turn leveling into a run of repeated exploits.[7]

3.1 The Shortcut Is Worth the Length of the Road

John Staats, whose development diary records the pathing saga, wrote the expectation down as a rule. Wherever players can exploit a game for experi-

ence, items, currency or reputation, they will, because they take the path of least resistance. Then the consequence: “if a game’s path of least resistance isn’t fun, it means the game isn’t fun.”[8]

The sentence puts the dull route on the designer’s side. Every reward can be reached by more than one route, each with a cost in minutes, risk and the walk back from a graveyard, and the designer sets all of those costs, including the ones nobody thought about. The intended route to a level is a long series of fights against monsters that fight back. The immunity spot reaches the same level with nothing fighting back, and the player pays in boredom alone. Chapter 15 gives the test for the word exploit: a route that collects the reward while skipping what the reward was meant to demand. The spot leaves the loot table intact and the fight absent.

Length decides how hard the dull route pulls. Content in a game of this kind is measured in months, not hours, “so any shortcut to the top will become the most popular route, even if it isn’t fun.”[9] The value of a shortcut grows with the road it skips: in a game that lasts an evening a dull trick saves minutes, and over months it saves weeks. Chapter 1 recorded players treating even travel time as downtime, and people who resent a long ride do not walk past a free level.

So the cheap response, addressed to players, fails. Warnings and punishments leave the route as it is and place the fault outside the design. Staats’s verdict on developers who blame players for aberrant behavior is that they are like dog owners blaming their pets for eating unhealthy scraps.[10] A player on the ledge is not a cheat. Games applaud clever use of terrain, and nothing marks which terrain the designer meant.

3.2 The Team Fixed the Route and Left the Player Alone

The first to find such routes were the team’s own colleagues, when the game opened to the rest of the company for an internal alpha in September 2002. Roman Kenney, an artist and *EverQuest* veteran, was known as the most creative exploit finder at Blizzard. Riding one of the game’s flying taxis, he dismounted in the air and fell into a high-level zone sealed off as unfinished. The fall killed him. He retrieved his corpse and found partly implemented vendors selling weapons better than anything the starting zone’s loot tables could produce.[11]

He linked the weapons in world chat, buyers who thought he had looted them made eager offers, and he resold at a huge profit. The designers were

amused at his ingenuity, though it sullied the economy, and they quickly removed dismount actions from the taxis.[12]

Kenney's second discovery was the campfire. Player-built campfires did damage, so he built them at the feet of characters whose owners were away from the keyboard, and the owners came back dead in a place meant to be safe. The designer Eric Dodds concluded that player campfires should do no damage. The reason they ever had: in the game's data, players were listed as an enemy of campfires.[13]

Neither fix involved a word to Kenney about how he ought to play. When outside testers arrived in November 2003 and hunted for holes as a sport, the episode Chapter 15 tells, exploits were reported and stamped out.[14]

The same response holds where nothing is broken. In the *Mists of Pandaria* expansion, Beastmastery hunters in arenas were stacking their cooldowns into one burst, the problem whose system-wide answer Chapter 9 follows.[15] Every ability did what its tooltip said. Greg Street's design post judged the burst not interesting, skillful or fair, blamed the team and not the players using the tools it had given them, and reduced it.[16] In both cases the designers read the player's discovery as a report on their own work.

Align optimal play with desirable play. A designer cannot ask players to take the route worth playing. The designer can only make it the route that pays.

The working idea is ***reward alignment***: adjusting what routes cost and return until the one a player picks for efficiency is the one the designer would pick for the experience. There are two levers. One raises the cost of the dull route, up to closing it. The other lowers the cost of the route worth playing. The remedy for immunity spots was the first kind, and it carried no message to players at all: better pathing code that left fewer spots.[17]

Magic: The Gathering reached the same principle from its own failure. Against gotcha, the *Unhinged* mechanic of Chapter 1, the correct defense was to stop talking, moving and laughing. Mark Rosewater's thirteenth lesson from twenty years of designing the game is to make the fun part of the game also the correct way to win. A game carries an implied promise that doing what it says will be enjoyable, and players who follow the instructions and are bored hold the game responsible.[18]

3.3 Pricing the Dull Route Out Before Anyone Takes It

The cleanest early repair belongs to Chapter 10: Gruul's Shatter, alone, would have made the best raid plan a seating chart, and a random knockback was added so that no chart survives the throw.[19] Assigned spots were never forbidden. The condition that made them work, knowing in advance where everyone would be, was removed.

The bounty system died on paper. Staats pitched it to Dodds: players would be assigned other players to hunt, like the afterschool game in which friends draw names and stalk one another with dart guns. Dodds played the feature in his head as its targets would. They could cheese it by staying offline, hiding in dungeons, sitting in zones where combat was disabled or surrounding themselves with guards. It also needed new tracking technology and duplicated a count the servers already kept. "Plus, searching wasn't particularly fun."[20]

Read as routes, a system meant to produce duels paid one side for hiding and the other for looking. Staats's next pitch, one-of-a-kind artifacts, failed because almost nobody on a server could ever get one. Where another studio's game relied on abstract incentives against griefing, this team simply removed the ability to grief.[21]

The third dull optimum involved no trick at all. In June 2003 Kevin Jordan judged that fights left too much downtime and felt too close to *EverQuest*, where one monster of similar level took most characters below half their health and players recovered by sitting still for minutes.[22]

The sitting was not laziness. When a second monster means death, the efficient player rests to full before every pull. The fix changed the price of a mistake. Jordan cut the damage monsters dealt so that they could be given pets and placed in pairs, and an extra pull would no longer kill the player. The cost was more monsters in the world and a risk to the frame rate.[23] Feedback from the August 2003 internal alpha said the overhaul was heading the right way, at least at low levels.[24]

> **Second-Order Effect: One arbitrage, one market**
>
> Kenney's fall could have stayed a private advantage: one character with a better sword. Because others could pay for the weapons, a hole in a zone boundary became a supply of goods, customers who misjudged their origin, and a profit.[25] The stakes named for immunity

> spots are the same: the integrity of the economy, ahead of any one player's unearned level.[26] In a persistent world with trade, a dominant route does not stay with the people who take it. It changes what a reward is worth for players who never went near it.

3.4 A Boring Optimum Built From Legal Cards

Magic has no walls to fall through. Its boring optimum is assembled from legal cards before the game starts, and the clearest recent case is energy, a mechanic Chapter 23 shows was judged too parasitic: each energy card wants other cards that make energy and others that spend it.[27] In January 2018, Ian Duke announced bans in Standard.[28] A deck called Temur Energy had begun the season at or above 20 percent of the competitive field and had reached about 30 percent of league play on *Magic Online*. The most played deck at the start of a Standard season has historically held about 10 percent. The start-of-season share was twice that norm, and the league share three times it, though from a later reading and a different population.[29]

The diagnosis was about price. The two banned cards each produced energy as a free extra on top of an effect nearly strong enough to justify the card alone, and they were the most common nonland cards in every version of the deck. It had stayed on top for five months, through a rotation, without enough counterplay.[30] A resource handed out at no cost made the deckbuilding question answer itself. Each energy card raised the value of the next, so the efficient list was the whole package.

The shape is older than energy: early poison, in Chapter 10, worked mostly with hard-to-block creatures, so one poison creature argued for a deck of them.[31] The standing guard against a strong card turning into a loop is the once-per-turn clause Chapter 4 examines on Rootwalla.[32]

The last case has no loop and no ban, and nothing in it was an oversight. *Odyssey*, from 2001, set out to challenge card advantage, the accepted principle that the player with more cards tends to win. Its threshold mechanic switched on with seven or more cards in the graveyard and turned Krosan Beast from a 1/1 into an 8/8. Patrol Hound's discard-for-first-strike was really a way to empty a hand into the graveyard. Throwing cards away was often the right play, as intended, and players disliked it: they wanted to play their cards.[33] The set performed poorly, and many of its discard effects offered a benefit that often did not matter.[34] Rosewater's lesson is

not to confuse interesting with fun.[35] The optimal strategy was chosen on purpose, and the people asked to execute it found it unpleasant.

3.5 Dominant, Degenerate, and Shipped Anyway

A ***dominant strategy*** does at least as well as every alternative whatever else happens. The immunity spot is the plain case, and resting to full when any second monster is fatal is another. A degenerate strategy, in Chapter 15's definition, is a legal, efficient way to play that makes the game worse for the person using it. Dominance is about payoffs, degeneracy about experience, and a dominant strategy that is a pleasure to carry out is a well-made game. The trouble starts when both words fit the same route and everyone has found it.

How soon they find it matters too. The design channel Distraction Makers names two diagnostics. A first-order optimal strategy is the first plan a player reaches and also the best, so later play has nothing more to give. Strategic collapse is a rule of thumb so strong that every other branch of play stops existing, and players blame the optimizers who exposed it when the fault sits with the designer who left it.[36] Fixed spots against a lone Shatter would have been first-order optimal. The energy package was a collapse in deckbuilding.

Reward alignment, the introduction's incentive compatibility, is the repair, and Table 3.1 sorts the chapter's fixes by lever.

Closing a route is certain and suits a route never meant to exist. Changing a payoff keeps the decision alive, the harder and usually better result: after the knockback the raid still chooses where to stand, and after the combat overhaul the player still chooses when to rest. Chapter 2 showed the same preference on one button, when raising the cost of Spellsteal turned a reflex into a judgment. The table has no row for leaving the route open and asking players to stay off it. A multiplayer *Magic* format has that row, a norm against bringing the strongest deck to a weaker table,[37] and Chapter 17 takes it up.

The principle has a limit. Designers sometimes ship the dull route on purpose, as with the potion chore of the introduction's Loatheb, accepted after internal debate for the final boss of a wing.[38] Closing a route costs decisions when the route was a real choice: Ulamog's Crusher, from Chapter 1, was forced to attack because playtesters kept it home, and whether to attack is the stuff the game is made of.[39] And a best route always exists. Duke's announcement says a recognized best deck is common in any season

Table 3.1. How each boring optimum in this chapter was handled, and the kind of fix each was.

The boring optimum	What the designers did	Kind of fix
Attack from an immunity spot	Rewrote the pathing	Closed the route
Dismount into a closed zone	Removed the dismount	Closed the route
Stack every cooldown	Reduced the burst	Changed the payoff
Stand on an assigned spot	Added a random throw	Changed the payoff
Rest after every fight	Cut monster damage, paired monsters	Changed the payoff
Play the energy package	Banned two cards	Closed part of the route

and is not itself a reason to ban, and energy decks were expected to stay viable.[40] The principle asks what the most efficient route is like to walk and what it has pushed out.

The Designer's Test

1. What is the cheapest route to this reward, counted in the player's minutes and risk, and is it one you would want to play?
2. If every player adopted that route, which part of your design would go unplayed?
3. Are you asking players to prefer the route you like, or have you priced it so that they will?
4. Does your fix change a payoff and keep the decision, or does it delete the decision?
5. Who besides the user pays when the cheap route is found: a buyer, an opponent, an economy?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section “Programming: The First Hurdle”.
2. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
3. Staats, *The WoW Diary*, section “Programming: The First Hurdle”.
4. Staats, *The WoW Diary*, section “April 2002: The Occasional Paradox”.
5. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
6. Staats, *The WoW Diary*, section “June 2003: A Crunchier Crunch”.
7. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
8. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
9. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
10. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
11. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
12. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
13. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
14. Staats, *The WoW Diary*, section “November 2003: Friends-and-Family Alpha”.
15. Arenas are the game’s format for ranked fights between small teams of players. A cooldown, as the introduction explained, is an ability that must recharge after each use.
16. Greg Street. Dev Watercooler: Mists of Pandaria PvP. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section “Beastmastery Hunters”.
17. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
18. Mark Rosewater. Twenty Years, Twenty Lessons, Part 2. Making Magic, `magic.wizards.com`, 2016, section “Lesson 13”.
19. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 0:37 and 0:56.
20. Staats, *The WoW Diary*, section “Announcement at the ECTS”.
21. Staats, *The WoW Diary*, section “Announcement at the ECTS”.
22. Staats, *The WoW Diary*, section “June 2003: A Crunchier Crunch”.
23. Staats, *The WoW Diary*, section “June 2003: A Crunchier Crunch”.
24. Staats, *The WoW Diary*, section “August 2003: Internal Alpha 4.0”.
25. Staats, *The WoW Diary*, section “September 2002: Internal Alpha 1.0”.
26. Staats, *The WoW Diary*, section “October 2002: Still Unanswered Questions”.
27. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section “Kaladesh, Lessons, Too parasitic”.
28. Standard is the tournament format built from the game’s most recent sets.
29. Ian Duke. January 15, 2018 Banned and Restricted Announcement. *`magic.wizards.com` (read from Wayback copy)*, 2018, section “Temur Energy and Other Energy Variants”.
30. Duke, section “Temur Energy and Other Energy Variants”.
31. Mark Rosewater. Something Wicked This Way Comes, Part 3. Making Magic, `magic.wizards.com`, 2010, section “Problem #3” and “Problem #4”.
32. Mark Rosewater. Why So Many Restrictions? Making Magic, `magic.wizards.com`, 2024, section “But Magic has been fun for many years”.
33. Mark Rosewater. Twenty Years, Twenty Lessons, Part 1. Making Magic, `magic.wizards.com`, 2016, section “Lesson 5”.
34. Mark Rosewater. Lessons Learned, Part 1. Making Magic, `magic.wizards.com`, 2023, section “Odyssey”.
35. Rosewater, “Twenty Lessons, Part 1”, section “Lesson 5”.
36. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 17, part B, at 11:19; episode 5 at 8:24; episode 20 at 15:42 and 17:25.

37. Distraction Makers, episode 23 at 4:49.
38. Candlesan, commentary, V6 at 2:26.
39. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section “Lesson 14”.
40. Duke, section “Temur Energy and Other Energy Variants”.

Chapter 4

Constraints Create the Game

The most anticipated item at the monthly team meetings on *World of Warcraft* was the design update, and in mid-2003 Allen Adham delivered it. He was soft-spoken, the room had to go quiet to hear him, and his delivery was so convincing that a team that usually argued rarely pushed back. That month he proposed removing cooldown timers from combat. Nothing would stop a player from casting the same spell again except a new system of diminishing returns on repeated actions. A weakened recast, he reasoned, was one more option than a forbidden one. Short casting times were cut to zero in the same pass.[1]

At first it felt better, on a visceral level, because nobody was waiting for a cooldown. Then the tests showed what was missing. Players needed "visual clues and hard stops to prevent them from losing all their mana."[2] People did not enjoy guessing how to be efficient, and after weeks of testing his combat models the approach was abandoned.[3]

The same run of tests produced the scene the team remembered. Encounters could now be built in an ability editor without a programmer, and timing, efficiency and strategy had begun to matter. A group of testers had just been wiped out by a hard encounter and were picking apart their tactics. Adham listened for a while, grinned, and asked the room whether they realized what was happening. "This is the first time ever when we have players talking about strategy. We just crossed a design milestone."[4]

The man who had argued for more options was the one who recognized what options worth discussing sound like: people arguing after a loss about what to do with the limited things they had.

4.1 A Price Leaves the Habit in Reach

Diminishing returns are the generous version of a common idea: keep every option open and steer with prices, so that nobody is forbidden anything and understanding the slope looks like skill. The idea fails for the reason Loatheb's two penalties failed in the introduction.

A graded penalty hands the player a price to estimate while the fight is running: how much weaker is the next cast, and is it still worth the mana. A hard stop hands the player a fact. The spell is ready or it is not, and that fact can be drawn on the button. The soft rule left casting again available, cheaper to reach than ever, and moved the cost into a mana pool that emptied quietly. The players who drained themselves were not careless. They answered the most visible thing on the screen, an enemy still standing and a spell that could still be cast. Adham's added option was the habit itself with the brakes removed.

A slope also has no moments. Each cast under diminishing returns is a slightly worse copy of the last, and nothing marks the point at which to stop and do something else. A cooldown manufactures that point. It ends one action cleanly and asks what comes next, which is why the testers had something to argue about. Nobody talks strategy about an action they can repeat freely. They talk about the cooldown spent too early.

The problem is general. A designer who wants players to cross a world on foot is up against their preference for arriving, and a card designer who wants a strong effect to be a decision is up against players using it as often as the rules allow. Each case has a reasonable default answer that makes the rest of the design irrelevant while it stays available. A design that leaves the habit in reach loses to the habit under pressure.

The difference between the two tools is the difference between a player's ***feasible set***, the actions available at all, and the payoffs attached to them:

> **Design equation (a conceptual framework, not a law)**
>
> | | |
> |---|---|
> | a price: | the same actions, with new payoffs |
> | a constraint: | fewer actions, with the old payoffs on those that remain |

Under a price the habit is still in the set, and the designer is betting the player will do the sum in time. Chapter 1 showed why that bet is poor when time is short. Under a constraint the player has to choose again from what is left, and choosing again is the activity the designer wanted.

4.2 A Constraint Takes an Answer and Leaves the Player

Games do not come from giving options. They come from restricting them in interesting ways.

The word doing the work is interesting. A rule is an ***interesting constraint*** when it does three things. It removes the habitual answer: the cooldown takes away the recast, where diminishing returns only made it expensive. It leaves more than one answer standing: a spell on cooldown leaves the question of when to spend it and what to do in the meantime. And it can be read without arithmetic: ready or not ready is a state, where a weakened recast is a sum.

The cooldown also changes what the action is worth. An action that can be taken once has ***option value***: it is worth more unspent than an action that can be repeated, because spending it gives up the chance to use it at a better moment. A spell that returns in two seconds has almost none. A spell that returns in a minute has a great deal, and all the arguing after a wipe is players trying to collect it. Scarcity did not make the action weaker. It gave it a second price, the moment forgone, and that price is what a player thinks about.

Legibility has a name inside the studio. The game's encounter designers use ***affordance***, a word borrowed from the study of everyday objects, for giving players the right amount of time and of visual and audio feedback to respond to a mechanic.[5] A hard stop affords in that sense. It has two states and an interface can display both. A slope has as many states as steps, and the player has to carry the table.

Two mechanics can be equally restrictive and differ in kind. When Akil'zon moved from a raid to a five-player dungeon, a conversion Chapter 13 follows, its old version threw a chosen player into the air and interrupted them, devastating when the pick was the healer or tank of a group with no spare. The new one has eagles carry the player off while the group kills the eagle, and the carried player can still act and cast.[6] Both pick a player and take something away. The difference refines the principle:

An interesting constraint takes away an answer. An uninteresting one takes away the player.

4.3 Two Factions, Fixed Routes and Bodies That Must Wear Armor

World of Warcraft made the same move at scales far larger than one spell. In October 2001 Adham resolved a question the team had been carrying: whether the game's good and evil races could play together. They could not. Orcs and humans would be unable to group or even communicate. He thought an "us and them" reinforced community and gave the races more personality, he wanted fighting between players confined to specific zones, as in *Dark Age of Camelot*, and splitting the population eased the fear of communities so large that nobody recognized anybody.[7] The rule reached the art department. Silhouettes became a major consideration in character design so that players could tell friend from enemy by shape alone, all the more once the population was split.[8]

The split passes all three tests. It removes a default, treating every other player as a possible ally and so as nobody in particular, it leaves a great deal standing, and it is legible at the distance of an outline on a hill.

Travel is the same move applied to the map. Rob Pardo explained at a 2006 keynote that the game used flight taxis, fixed routes between fixed points, where it could have offered free teleportation, and left some areas with no flight path at all, to make the world feel epic and to build social connection. Players called the game World of Walkcraft, and he defended the trade as what such games are about.[9] The counter-experiment came from the same studio. In *Cataclysm* the new high-level zones were scattered and joined by portals, players teleported to nearly every destination, and the expansion felt disjointed. The designers concluded they had underestimated how much a land to explore matters and planned the next expansion around a single continent.[10] Nobody forced a player through a portal. The habit, left available, won.

The roster was constrained twice. The team chose fewer classes, folding abilities from several *Warcraft III* classes into one, a result Pardo called concentrated coolness.[11] The other limit nobody chose: every race has to wear every piece of armor. Fitting helmets over tauren horns was hard enough,[12] and when nobody could give the naga a body that took armor and common animations, they stayed monsters, the case Chapter 22 counts in full.[13]

Some limits are not real. Dana Jan's Deadmines ended with a warship that could not be cleanly cut off from the outside world, so he enclosed it in a cavern bigger than interiors were supposed to be. Scott Hartin extended the engine's farclip, the distance a player can see before geometry disappears,

and warned that level designers would now build ever bigger rooms with no way to know what it would do to the frame rate. They did, and testing showed the cost was negligible,[14] a lesson later counted valuable for building epic areas.[15] An untested limit is a guess, and guesses shape games as firmly as facts until someone builds past them.

Second-Order Effect: The side that got less

The faction split gave the races personality, and it gave the team two sides to furnish. In October 2004 the artists were making Horde-specific outfits and weapons because the team had realized late in the project that more quests, outfits and buildings had been designed around the good guys, leaving Horde content thin.[16] A second cost came from abroad. During the beta, churn in Korea ran at 25 percent against 5 percent in North America, five times higher, and the team read the Korean gaming press as saying players there did not enjoy the Horde's monster races.[17] A wall between two populations means every piece of content is owed twice, and half the game can be the half a market does not want.

4.4 Once Per Turn

Magic: The Gathering runs on the same principle at the size of a card. Mark Rosewater's lesson is that restrictions breed creativity: a designer given complete freedom reuses familiar solutions, and a new restriction forces the search to begin elsewhere. Its companion asks how little needs to be added. A multicolor block that had to avoid repeating *Invasion* was given one rule, exactly two colors at a time with all ten pairs treated equally as guilds, and players embraced the guilds without confusing the two blocks.[18]

Beneath every set lies a constraint players stop seeing. A card has a color, casting it takes mana of that color, and each color is good at some things and bad at others, the allocation called the color pie.[19] A deck that wants every strong effect pays in the difficulty of producing every color.

What that constraint is worth shows where it was loosened. Artifacts are usually colorless, so any deck can cast them. In *Mirrodin*, a block built to push artifacts as far as they would go, Rosewater pressed for colored mana in artifact costs, gave way when a colleague asked for most of it to be removed, and now regards giving way as a mistake. The block had serious problems in play because every deck could use every problematic card.[20]

The problem returned in the *Kaladesh* block and went unsolved there.[21] It is *Cataclysm*'s portals in cardboard: remove the friction that kept things apart, and everything arrives everywhere at once.

On a single card the restriction is part of the design, because a designer who controls when and how often an effect happens can decide how large it may be. The standard example is Rootwalla, a creature that may make itself bigger once per turn when similar abilities elsewhere could be used repeatedly, and its name became studio shorthand for the pattern. A timing rule from the game's origins works the same way: most cards can be played only in the player's own main phase, because Richard, who built many of the early restrictions, wanted a default for how a turn proceeds, with instants as the exception.[22] Once per turn is a cooldown on cardboard. The limit is what lets the effect be strong.

The pattern runs in both directions. Kicker let a player pay extra for a bigger spell and grew until it meant paying extra to do nearly anything, at which point players greeted new mechanics as just kicker. It was too flexible, and the repair narrowed it to two outcomes, a larger effect or a second related one.[23] Investigate, built for a set about mysteries, first drew a card outright, which put too much card drawing in the set. The shipped version makes a Clue, a token its owner turns into a card later by spending mana.[24] A reward handed over at once became one with option value, scheduled against everything else the mana could buy. The designers bind themselves the same way: early choices about a set's vision cut away large regions of the possible and save the effort that would be wasted there.[25]

4.5 The Queen Problem

The strongest objection to this chapter comes from the design channel Distraction Makers. Its Queen problem asks what stops everyone from bringing the strongest pieces once players pick their own. The answers sort into three: a restriction (a piece may or may not be used), a limitation (how many of it) and a resource, a cost paid during play. The hosts prefer the resource. Rules imposed from outside play feel arbitrary and pile up, while under a cost nothing is forbidden and the strong piece simply does not work yet.[26]

That reverses the ranking above and adds a middle term the two lines of the framework lack. The reply is partial. At the table an unpaid mana cost reads as a hard stop, a card that can be cast or cannot, so the resource they prefer works because it is legible in the same way a cooldown is. The objection also has evidence behind it. The studio has become blunter about

restrictions, stating outright limits once folded quietly into other features of a card, and players find some of the new wording unpleasant to read.[27] The clearest way to state a limit is also the one that reads most like a prohibition. And a constraint can be aimed at the wrong thing: Chapter 15 tells how private auras, a limit on what players could know, sometimes proved worse than the add-on problem they were meant to solve.[28]

The Designer's Test

1. Which habitual answer does this constraint remove? If you cannot name it, the constraint is removing something else.
2. After it is removed, how many answers are left, and can the player see them without doing a sum?
3. Does the constraint take away an option, or the player's ability to act?
4. Is the thing you limited worth more now that it is scarce, and is there a moment when the player has to decide to spend it?
5. What does the constraint force you to build twice, and is the limit real or has nobody tested it?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "June 2003: A Crunchier Crunch".
2. Mana is the pool a caster spends to cast spells. A caster with none left has, for practical purposes, stopped playing until it refills.
3. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
4. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
5. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, affordance answer.
6. Scott Mercer. Blizzard Interview with Lead Encounter Designer Scott Mercer. Wowhead, 2011, Akil'zon answer.
7. Staats, *The WoW Diary*, section "October 2001: Learning from the Good and Bad".
8. Staats, *The WoW Diary*, section "Character Design".
9. Mark Wallace. Report of Rob Pardo's keynote on World of Warcraft at the Austin Game Conference. Game Developer, 2006, paragraph 10.
10. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What didn't work out as planned?".
11. Wallace, report of Pardo's Austin keynote, paragraph 6.
12. Staats, *The WoW Diary*, section "Character Design".

13. Staats, *The WoW Diary*, section “November 2002: Internal Alpha 2.0”.
14. Staats, *The WoW Diary*, section “April 2002: The Occasional Paradox”.
15. Staats, *The WoW Diary*, section “December 2003: Stepping on Toes”.
16. Staats, *The WoW Diary*, section “October 2004: World’s End”.
17. Staats, *The WoW Diary*, section “July 2004: Public Beta 3.0”.
18. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, sections “Lesson 17” and “Lesson 18”.
19. Mana in *Magic* is the resource a player spends to cast cards. The color pie is the designers’ division of effects among the game’s colors, so that no single color can do everything.
20. Mark Rosewater. Lessons Learned, Part 1. Making Magic, `magic.wizards.com`, 2023, section “Mirrodin”.
21. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section “Kaladesh, Lessons”.
22. Mark Rosewater. Why So Many Restrictions? Making Magic, `magic.wizards.com`, 2024, sections “Why even have restrictions?”, “Rootwalla” and “Sorcery Speed”.
23. Mark Rosewater. Here’s the Kicker. Making Magic, `magic.wizards.com`, 2007, section “Kicked When You’re Down”.
24. Mark Rosewater. Looking Back, Part 2. Making Magic, `magic.wizards.com`, 2024, section “2016 Best Addition”.
25. Mark Rosewater. Nuts & Bolts 7: The Three Stages of Design. Making Magic, `magic.wizards.com`, 2015, section “The Vision Stage”.
26. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 17, part B, at 0:09, 1:44 and 5:49.
27. Rosewater, “Why So Many Restrictions?”, section “Does it bother you that some of the new restrictions feel bad to read?”.
28. De Sousa and Sanders, interview, private auras answer.

Part II

Uncertainty

Chapter 5

A Solved Game Stops Being a Game

In October 2004, a month before *World of Warcraft* reached store shelves, its designers put forty or so developers into the Molten Core, a raid built for forty players fighting together. Getting forty people to test as a group was hard, and it was the only way to balance the combat.[1]

The session went badly. The designers described each fight in chat and were ignored. There was no voice software and no raid interface, so a player could see health bars only for the few members of an immediate party. Jeff Kaplan, the designer running the numbers, had no internal tool for measuring the group's damage and used meters fans had written as add-ons during the beta. The volunteers wiped on every pull of the trash, the ordinary monsters that stand between a raid and its bosses, and never reached a boss.[2] Many developers doubted players would ever kill these monsters. Kaplan typed that organized raids would get past the content easily.[3]

In later sessions the group reached its first boss, a pack of lava hounds, and still lost half the raid on each pull. Kaplan explained that the hounds all had to die within seconds of one another. Many developers cheered, and Kaplan chose a stronger word: "Trust me, once the raiders figure this out, it will be trivial."[4]

The open beta in November did no better. The first trash monster wiped group after group of players in poor armor with no buffs or flasks. Fans concluded the content was bugged or guessed that the level cap would rise. Inside the team raids looked likely to appeal to 1 percent of the player base, one player in a hundred. Kaplan stuck to his math.[5]

5.1 Knowledge Turns a Wall Into a Routine

Between the impossible fight and Kaplan's trivial one, nothing in the fight would change except what the players knew. If knowledge alone moves an

encounter from wall to routine, the encounter is a game only in the interval between them, while a raid is still working out what is being asked of it.

The lava hounds show how short the interval is. What was killing the developers had an answer one sentence long, and Kaplan supplied it in chat. A sentence like that does not stay with the people who earned it. The team accepted as much. When a task force of the forty best players in QA still could not find every bug, the designers opened a public test server: "having spoilers on a test server was far better than broken boss fights."[6] The designer's question cannot be how to keep a fight unknown. It is what the fight still asks of a raid that arrives knowing.

The tidy answer is a fair puzzle learned completely: fixed abilities, fixed timings, the same order every time. The game's two current lead encounter designers describe Gallywix, a recent final boss, as a fight whose mechanics are assigned to players in advance, a fight a team improves at through planning and collaboration.[7] A fixed fight is also one a tuner can measure, which is what Kaplan was doing with screenshots of damage output and combat logs after every wipe.[8]

It gets solved before the pull. Gruul's Shatter, which Chapter 10 follows, would alone have been answered by a seating chart.[9] A mechanic that can be answered on a diagram has been answered, and the fight becomes a recital of the diagram. This is the sense in which a game, or a piece of one, is ***solved***: the best response is known before play begins and does not depend on anything that happens during it. (Chapter 10 uses the narrower word precomputed for a single mechanic.) A solved fight is a dominant strategy, in Chapter 3's sense, whose discovery is finished and published. The designers who praised the pre-assigned fight set it beside fights that make a raid think on its feet and said a great raid has both.[10]

5.2 Players Buy Certainty Wherever It Is Sold

Prophet Skeram splits into copies whose only intended tell is health: they have half of it, so their bars fall faster. Testers found interface mods that read creature IDs, and after release a hunter's pet sent at the boss before the split stayed locked on the real one, the contest Chapter 15 tells.[11] The design offered a judgment that took seconds and could be wrong. Players replaced it with a lookup that took none and could not. At the top of the game the lookup became an industry, with race teams hiring programmers to write add-ons that computed mechanics, and the difficulty that was left

settled on dexterity and reflex.[12] In the October test it was fans, not the studio, who had already built the meters.

None of this is misbehavior. The tool that explains it is the ***value of information***: the most a decision maker should pay to learn a fact before choosing, equal to how much better the decision is expected to turn out with the fact than without it, and zero when the fact would not change the choice. Push the real Skeram to his next split with two illusions alive and there are five Skerams, and the raid wipes.[13] A tell that names the real one is worth close to a whole attempt, for forty people, every time. The guides, meters and hired programmers are each worth exactly the uncertainty they remove.

The same sum gives the designer a test. Information that is cheap, available outside the fight and decisive will be bought, and the decision it resolves is gone. Information that must be gathered inside the fight, at some cost in time or risk, is a decision in its own right. Watching health bars fall was meant to be the second kind. A creature ID is the first.

5.3 Eight Sources of Doubt

A game is alive while the player still faces something that cannot be settled before acting. Players will remove every uncertainty they can, so the designer's real choice is which uncertainty survives them.

Whether performing a script is still a game depends on whether the performance is in doubt, the subject of Chapter 8. To say which doubt a mechanic holds, the book borrows a vocabulary. Greg Costikyan's *Uncertainty in Games* names the sources from which games draw uncertainty, and eight of his names serve from here on as a toolbox, sorted into two groups of four.[14]

Four concern what is in front of the player. ***Randomness*** is an outcome drawn fresh: Gruul's Ground Slam throws each raider a random distance in a random direction, and a shuffle orders a *Magic* deck.[15] ***Hidden information*** is a fact that exists and is withheld, such as which Skeram is real or which cards an opponent holds. ***Player unpredictability*** is doubt about what another person will choose, across a card table the opponent, in a raid every ally whose next move a player cannot see. ***Performative uncertainty*** is doubt about whether a player can carry out with hands and eyes what the player already knows. Knowing the lava hounds must die together is one thing. Forty people making them die together is another.

Four concern what the player has not yet worked out or been given. ***Analytic complexity*** is everything in plain view with too many consequences to think through: every *Magic* card prints its full text, and which cards belong in a deck stays open. ***Solver uncertainty*** is the doubt inside a puzzle whose answer the designer is holding, the developers' position before the hounds until Kaplan spoke. ***Narrative anticipation*** is wanting to learn how a story comes out, as in any quest line whose last step the player has not reached. ***Development anticipation*** is doubt about what the makers will do next, the beta fans guessing at a higher level cap (Figure 5.1).

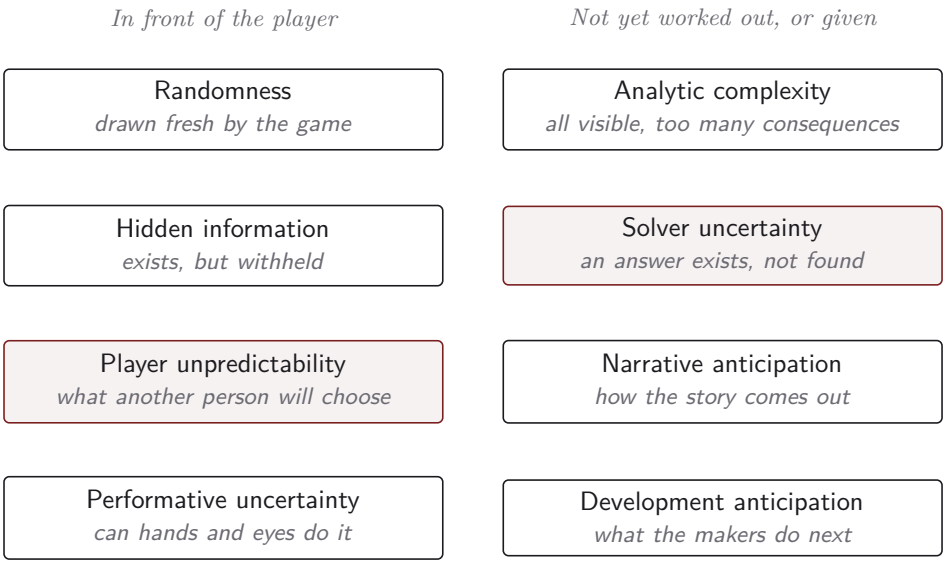


Figure 5.1. Eight of Greg Costikyan's sources of uncertainty, in two groups of four. Accented are the two that stopped the volunteers in the Molten Core.

Two questions follow, and the book keeps asking them: which uncertainty does this mechanic create, and is it the one the designer wanted? In the Molten Core the volunteers were stopped by solver uncertainty (nobody knew the trick) and by player unpredictability among allies (nobody could see or hear the rest of the raid). If solver uncertainty is all an encounter has, one sentence in chat ends it. A seating chart is that sentence for a lone Shatter, draining out the uncertainty its designer wanted: who ends up where, and whether each player gets clear in time. Skeram was built on hidden information, and the game's own interface leaked it.

5.4 Putting the Unknown Back, or Never Letting It Out

The first repair moves the unknown to a moment study cannot reach. The four Anubisath Sentinels at the entrance to Skeram's raid each pick one ability from a pool of eight when the group is pulled, so the raid must read what it faces, reposition and choose a kill order on the spot, the case Chapter 12 tells.[16] The pick falls after a raid's last chance to plan and before its first decision. A guide can list all eight abilities. It cannot say which ones stand at the door tonight. Study still pays but no longer finishes the job. The current encounter leads keep the same device, random assignments in the fights Stix and Sprocket, beside the pre-assigned mechanics of Gallywix.[17]

The second repair recombines what players already learned. On Mythic, the hardest difficulty, the leads reorganized Gallywix's phases and reused the encounter's elements in a flipped order, so that players who knew the fight had to recognize a new approach at once.[18] That is solver uncertainty rebuilt from used parts, at the cost of no new mechanics. In hindsight Taylor Sanders judged it unprecedented in modern WoW for the penultimate and final bosses to sit so close in difficulty, and a wider gap would have been better.[19] Novelty and difficulty are separate dials.

Space can be solved as well as fights. In August 2003 the same ice cave sat four times in the starting zone of Dun Morogh, and the producers assured John Staats, the level designer whose diary records it, that nobody cared about repeated layouts.[20] Then at a weekly board-game night the tech support manager, John Hsieh, admitted he had stopped entering those caves during the latest internal alpha once he realized they were one layout, and others agreed. Staats built variations on his own weekends, and once they plainly improved exploration a producer had every duplicate in the world listed. "Instead of having three caves and two goldmines, we had fifteen variations of each."[21] A cave holds a small piece of hidden information, what lies around the next bend. The second copy holds none. Hsieh had solved it before walking in.

The last repair broke a certainty the designers had built themselves. Smart heals choose their own target, and they had always picked the most injured ally. A design post changed them to pick at random among injured allies in range, so that healers would care whom they targeted and which heal they used.[22] A helper that always finds the best target answers the question the healer was there to answer, and randomness made the person's judgment needed again.

Second-Order Effect: The mines nobody finished

The cave variations restored exploration and then overshot. Some were far too big. The Mulgore mines ran so long that players passed fifth level before reaching the end, and Staats has yet to meet anyone who played through his goldmines in the Blasted Lands.[23] Unknown space is worth building only if somebody arrives to resolve it. A cave the player outgrows halfway through is as wasted as the fourth copy of one already seen.

Magic: The Gathering does not have to put uncertainty back, because two kinds are written into the rulebook where no player can reach them. The Comprehensive Rules sort every zone into public and hidden. The library, the deck a player draws from, is kept face down and no player may look at or change its order. A player may count the cards in an opponent's hand and may not look at them.[24] Draw order is randomness, the opponent's hand is hidden information, and no add-on dissolves either.

Mark Rosewater's checklist of what a set structurally needs includes tools that offset the draw, such as scry, surveil and ordinary card drawing.[25] He calls the mana system one of the game's greatest sources of variance and defends it: it keeps games different, lets the weaker player sometimes win, and opens a second area of skill in building a deck's mana and coping when it goes wrong.[26] Chapter 6 takes up what that variance costs the player who suffers it.

Two more sit outside any single game: the field of decks opponents bring, the metagame, and the player's own deck construction, analytic complexity worked against the other three. The metagame will not hold still. A card's usefulness can change sharply as the decks around it change, a spread of card power tilts results toward skill, and R&D cannot make good cards that look bad without also making bad cards that are bad.[27] A deck solved against this month's field is a wager about next month's, because the answer changes the question once enough players adopt it.

The doubt also depends on who is looking. Beginners are not troubled by strategic complexity because they cannot yet see it, the distinction Chapter 21 builds on.[28] In the Molten Core knowledge closed the game, and in *Magic* it opens it: a card that is finished business for a beginner is a live question for an expert. And the clock resets. The first thing players ask of a new set is what is new, and the format built from freshly opened cards

should encourage discovery.[29][30] The release schedule is analytic complexity delivered on a calendar.

5.5 When Solved Is Not Dead

Kaplan's word was a promise. Mastery that turns hard content into routine is what raiders work toward, and a boss on farm is an achievement.[31] Kaplan held his position through the open beta, and other experienced raiders agreed the content was exactly as hard as it needed to be.[32] A solved fight that forty people still have to execute is not empty. A solved game stops being the game it was and may become a test of execution or a reward, both of which players want.

The opposite error is uncertainty the player cannot resolve. Early versions of one *Magic* set rewarded having poison counters, players could not tell whether to attack with their poison creatures, and the cards were removed, the case Chapter 9 tells.[33] A guess is not a decision. An unknown earns its place only when playing well can reduce it.

And solved, for a community, is a belief before it is a fact. A crowd that sees the winning decks, copies them and stops experimenting makes those decks look strongest partly because nothing else is tried, and a field players believe is settled stalls as a settled one would.[34] An uncertainty can survive in the game and be gone from the players' heads.

The Designer's Test

1. Can this problem be solved before play begins?
2. Once it is solved, what is the player still uncertain about?
3. Which kind of uncertainty is that, and is it the kind you meant to create?
4. Does the player have any way to reduce it by playing well, or only by guessing?
5. What will players build, write or share to remove it, and what is left after they do?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "October 2004: World's End".
2. To pull is to start a fight by drawing the enemies' attention. A wipe is a failed attempt in which the whole group dies and starts over.
3. Staats, *The WoW Diary*, section "October 2004: World's End".
4. Staats, *The WoW Diary*, section "October 2004: World's End".
5. Staats, *The WoW Diary*, section "November 2004: Open Beta".
6. Staats, *The WoW Diary*, section "October 2004: World's End".
7. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "One of the things that we had between different bosses".
8. Staats, *The WoW Diary*, section "October 2004: World's End".
9. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 0:37.
10. De Sousa and Sanders, interview, answer to "One of the things that we had between different bosses".
11. Candlesan, commentary, V3 at 1:31 to 2:47.
12. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section "ADDONS & THE TRICKLE DOWN EFFECT".
13. Candlesan, commentary, V3 at 1:51.
14. Greg Costikyan. *Uncertainty in Games*. The MIT Press (Playful Thinking series), 2015.
15. Candlesan, commentary, V2 at 0:59.
16. Candlesan, commentary, V5 at 2:43.
17. De Sousa and Sanders, interview, answer to "One of the things that we had between different bosses".
18. De Sousa and Sanders, interview, answer beginning "We saw Mythic Gallywix had a completely different design".
19. De Sousa and Sanders, interview, answer to "How do you view Mythic Gallywix's overall difficulty in hindsight?".
20. Staats, *The WoW Diary*, section "August 2003: Internal Alpha 4.0".
21. Staats, *The WoW Diary*, section "September 2003: A Sense of Place".
22. Blizzard Entertainment. Dev Watercooler: Building Healthy Gameplay. `worldofwarcraft.blizzard.com`, n.d., section "Retuning Healing Spells".
23. Staats, *The WoW Diary*, section "September 2003: A Sense of Place".
24. Wizards of the Coast. Magic: The Gathering Comprehensive Rules (effective January 16, 2026). `media.wizards.com`, 2026, rules 400.2, 401.2 and 402.3.
25. Mark Rosewater. Nuts & Bolts 15: Structural Support. Making Magic, `magic.wizards.com`, 2023, section "Deck Smoothing".
26. Mark Rosewater. Mana Action. Making Magic, `magic.wizards.com`, 2011, sections "#6" and "#7".
27. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, sections 4 to 6.
28. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "Common Knowledge".
29. Mark Rosewater. Nuts & Bolts 9: Evaluation. Making Magic, `magic.wizards.com`, 2017, section "Question #6".
30. Mark Rosewater. Nuts & Bolts 2: Design Skeleton. Making Magic, `magic.wizards.com`, 2010, section "#1".

31. A boss is on farm when a raid kills it routinely, week after week, for its loot.
32. Staats, *The WoW Diary*, section “November 2004: Open Beta”.
33. Rosewater, “Evaluation”, section “Question #3”.
34. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 7 at 13:33 and 14:06.

Chapter 6

Random Is Not the Same as Fair

An ability has a 25 percent chance to land a critical hit. A player uses it four times and gets four ordinary hits. That is disappointing and nothing more. The player uses it four more times and gets four more ordinary hits, and by the eighth swing the disappointment has turned into a different thought. The player has stopped thinking about luck and started thinking about the game. Something, the player decides, is wrong with it.

Nothing is. With a 25 percent crit rate, four swings in a row produce no crit 31.6 percent of the time, about one sequence in three, and eight swings do so 10 percent of the time, one in ten.[1] The generator is doing what it was told, and no audit would fail it. Candlesan, the encounter designer whose example this is, gives the verdict in one line: “Your math is right, but your game is still going to feel broken.”[2] Every roll is stateless, so nothing stops a bad streak, and bad streaks make players quit.[3]

His repair is an object a child could build, a marble bag. For a 25 percent chance, put two red marbles and six blue ones in a bag. Each time the game needs an outcome, draw a marble and do not put it back. When the bag is empty, refill it and carry on.[4] Over any full bag the rate is two in eight, the same 25 percent as before. What has changed is that the draws now depend on one another. Six blue marbles in a row from a full bag leave only red ones behind. In his words, “we’ve given the randomness a memory.”[5]

6.1 A Correct Generator Builds Its Own Case Against Itself

The designer promised a rate. The player lives through a sequence. Twenty-five percent describes thousands of swings and says almost nothing about the next eight, and a session of play is the next eight. The player remembers

the worst stretch and holds it against the tooltip. Both readings are correct, of different objects.

The engineer's answer is to keep the generator correct and prove it: publish the rate, let players log ten thousand swings. It is honest and easy to verify, and a designer who bends the dice has started deciding outcomes for the player. But the streak the player calls impossible is close to guaranteed. If each swing crits with probability p, independently of the others, the chance that n swings in a row all fail is

$$P(\text{no crit in } n) = (1 - p)^n.$$

At $p = 0.25$ that gives the 31.6 percent for four swings and the 10.0 percent for eight. A long fight does not test one window. Every swing begins a new window of eight, and the book's script counts the overlapping windows. Across 100 independent swings, a run of eight straight misses appears somewhere 95.7 percent of the time. Against 10.0 percent for any one window, that is the difference between an accident and an appointment. A run of twelve, 3.2 percent in a single window, about one in thirty, appears somewhere in the hundred 55.4 percent of the time, better than a coin flip.

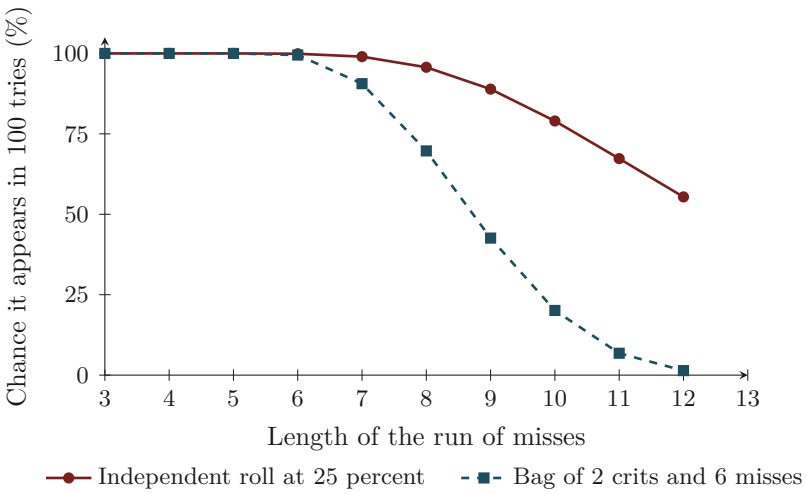


Figure 6.1. The chance that a run of consecutive misses of a given length appears at least once in 100 tries, for a stateless 25 percent roll and for a marble bag of two crits and six misses refilled when empty. Under the bag a run of 13 cannot occur. Author's calculation.

The player's sense that eight misses should be unusual is right about the window and wrong about the evening. A correct generator, left alone, manufactures the evidence that it is broken, and it does so on schedule.

The stateless rolls are independent ***Bernoulli trials***: after seven misses the chance of an eighth is still three in four, and the wait for the next crit is ***memoryless***. People do not judge sequences that way. They judge a short run by how much it resembles the process behind it, the habit Amos Tversky and Daniel Kahneman named the representativeness heuristic,[6] and they expect independent events to even out over short stretches.[7] The gambler's fallacy travels with it: a streak makes the opposite outcome due. When the stretch fails to even out, somebody is responsible, and the dice have no author.

Publishing the rate makes it worse, since a stated 25 percent is a promise the player can measure the last eight swings against. A team may be able to prove its game balanced, and the proof does not matter if players do not think so.[8] A proof addresses the rate. The complaint is about the sequence.

6.2 Memory Is a Dial, and Someone Can Read It

Random is not the same as fair. The sentence is Candlesan's closing line.[9] For the person playing, fairness is a property of the sequence they saw, and the designer sets it with three choices, of which the rate is the least.

The first is whether the randomness has memory. A stateless roll guarantees its rate only in the limit. A bag guarantees it inside every refill. In a bag the gambler's fallacy is true by construction: after six blue marbles a red one really is owed. The designer built a world that matches the player's error.

The second is how much memory. Figure 6.1 shows what a bag of eight buys. The longest dry run it can produce is twelve, the six blue marbles that end one bag followed by the six that begin the next, and a run of thirteen is impossible, where the stateless roll has no ceiling at all. A run of twelve somewhere in 100 draws falls from 55.4 percent to 1.4 percent. A run of eight still appears 69.7 percent of the time, down from 95.7. The bag removes the catastrophe and keeps most of the ordinary bad luck. A bag of one red and three blue holds the same rate and cannot run dry past six draws. Candlesan also nests bags, with one marble in the outer bag leading to an inner bag of rare outcomes, which keeps them rare and spread apart.[10] His cascading loot roll, which Chapter 18 takes apart, fixes first the probability players care about most.[11] The size and nesting of the bag are dials.

The third is who can read the memory. A bag whose contents are hidden corrects streaks quietly. A countable bag hands the player a forecast: once both red marbles are out, every draw left is blue. Memory turns part of the randomness into hidden information (Chapter 7).

World of Warcraft has turned all three dials, and its trinkets turned them in sequence. Before patch 5.2, such effects had a high chance to trigger followed by a fixed lockout, which made them random in theory and predictable in practice: once at the start of a fight, then at a steady interval. Patch 5.2 set how often an effect should trigger per minute and left the timing to chance, and then a trinket meant to trigger about once a minute could go eight minutes or more without doing so.[12] The hotfix tracks the time since the last success, and once that passes one and a half times the average interval it multiplies the chance by a factor that grows with the wait: 1.3 at 72 seconds against a 45-second average, 16.5 after five minutes. Unlucky streaks became rarer and the extreme ones disappeared.[13] Memory so strong it was a timer, then no memory and eight-minute droughts, then a little memory applied only in the tail, where Figure 6.1 puts the damage.

6.3 The Price of a Failed Roll

The game did not come to randomness reluctantly. At a design meeting in October 2003, John Staats records, monster loot felt too repetitive and limited, and the team borrowed semi-random item generation from *Diablo*.[14] Seven months later the beta had empty dungeons and missing quests, and testers stayed: "The drive for randomly acquired items and socializing kept testers happy and occupied."[15]

The same engine misfired when each roll grew expensive. Archaeology, the *Cataclysm* profession built on random rewards, long-term goals and a great deal of travel, was received as grindy, too random and time-consuming, and the plan for the next expansion was less travel and less randomness.[16] A failed crit costs a second. A failed dig cost a trip. Variance that is tolerable at one price per trial stops being tolerable at many times that price.

Blizzard's later answers to expensive failures give the randomness a memory in different shapes. For a holiday boss, a 2024 event notice says the chance of the sought item rises with each failed daily attempt and keeps accumulating, so that players spend less time chasing one drop.[17] That is the marble bag with a slope. For one raid season in 2022, a currency capped at three bought a chosen item outright, covering extreme bad luck without undercutting gear earned with a group.[18] Memory is not automatically kind,

though: the *Legion* legendaries remembered a player's luck against them until a 2017 repair (Chapter 18).[19]

When the roll falls matters as much as what it costs. Automatically targeted heals were changed to pick at random among injured allies, so that healers would have to care about targeting.[20] Gruul's Ground Slam throws each raider a random distance in a random direction and then asks for a response (Chapter 10), and each Anubisath Sentinel draws its ability at the pull for the raid to read (Chapter 12). In all three the roll comes before the decision. A crit or a drop comes after it, and the argument about fairness gathers around that second kind, where the dice have the last word. Input randomness falls before the critical decision, as a drawn hand does, and output randomness after it, and players feel the second far more even when the arithmetic is equal. A roll before the last act becomes the input to the next decision. A roll at the very end leaves no turn to answer it, and players see it as having decided the game.[21]

6.4 A Deck Is a Bag Sampled Thinly

Mechanically, the marble bag is a deck of cards: two red, six blue, shuffle, deal off the top, shuffle again when empty.[22] *Magic: The Gathering* is played from a shuffled deck, and a drawn card does not go back. If memory cured the feeling of unfairness, its players would have nothing to complain about.

They complain anyway. Lands produce mana, the resource every other card costs, and a hand with too few of them cannot act. Mark Rosewater's defense of the mana system (Chapter 5) calls it one of the game's greatest sources of randomness and variance, and mana screw a by-product of a system worth having.[23] Table 6.1 treats two to four lands in a seven-card hand as workable, a simplification and not a rule.

Lands in the deck	0 or 1 land	2 to 4 lands	5 or more
20	24.7%	71.7%	3.6%
22	19.0%	75.4%	5.6%
24	14.3%	77.5%	8.3%
26	10.4%	77.8%	11.8%

Table 6.1. Lands in a seven-card opening hand from a 60-card deck, by the number of lands in the deck. Author's calculation.

With 24 lands, 14.3 percent of opening hands hold one land or none, about one hand in seven, and 8.3 percent hold five or more, about one in

twelve. Adding lands does not fix this. Going from 20 to 26 cuts the starved hands from 24.7 percent to 10.4 and more than triples the flooded ones, from 3.6 percent to 11.8, while the middle column moves by six points. The marble bag is emptied completely. A hand is seven cards out of sixty, and a bag sampled this thinly keeps nearly all the variance of an independent roll. Memory helps in proportion to how much of the bag the player gets through.

The designers answer at three levels. Inside a set, Rosewater's structural checklist asks for one mechanic that does most of the draw smoothing, things to spend surplus mana on late, and curves that do not clump at the commonest rarity.[24]

In the rules, the answer is the mulligan, the right to throw back an opening hand, and its history is a price being revised. In 2015 a player shuffled the hand away and drew one card fewer, and a Pro Tour trial added scry 1, a small selection over the cards to come, because too many games had no interaction after unfortunate mulligans.[25] In 2019 the London mulligan replaced the rule in every format: draw a full seven each time, then put one card on the bottom for each mulligan taken. The aim was more competitive games when players had mulliganed unequal numbers of times, the testing watched for decks assembling the same opening every game, and a player with a reasonable hand should still prefer to keep it.[26] Twice in four years the cost of a bad draw was lowered, and both times a cost was kept.

The digital game filters in secret. In single-game matches of *Magic: The Gathering Arena*, the system drew an opening hand from each of two separately shuffled copies of the deck and leaned toward the one whose mix of lands and spells was closer to the deck's average, because a mulligan felt worse with no further games to recover in, while not rewarding unusual mana bases.[27] The patch of 14 February 2019 raised the pool to three shuffles and extended the same filter to mulligan hands.[28] It is a filter on the tails, applied where the player cannot see it.

6.5 Some Players Came for the Spike

Smoothing has a floor, and it is certainty. *Cataclysm*'s legendary weapon could be completed with certainty, guilds came to expect legendaries in turn, and the result felt like entitlement. A legendary should be rare and exciting, not a bar to be filled, and the plan was to make them rarer.[29] A guaranteed reward is treated as a wage, and a late wage is a grievance. The 2022 currency's cap of three marks that limit.

Nor do all players want the same thing from chance. Molten Sentry is a *Magic* creature built on a coin flip between two balanced results. It pleased nobody: those who enjoy variance found the outcomes too even to feel like a gamble, and those who resent randomness disliked a coin overriding their play.[30] Chapter 14 draws the lesson about audiences. Variation in card strength also tilts results toward skill, because telling a good card from a bad one is something players can be better or worse at,[31] so removing variance to protect the better player can remove what the better player was good at reading.

Luck also serves the loser. A player who has lost a match commonly says the draw was unlucky, and the claim hides the earlier decision that led to the loss, which is why players who turn competitive learn to stop saying it. It also spares the player the verdict of not being good enough, a release valve that a game with no luck lacks.[32] A designer who smooths the draw takes away that alibi along with the streak. And since the bag, the rising drop chance and the filtered draw all depart from the rule the player believes is in force, a designer who gives the dice a memory also decides who will be told.

The Designer's Test

1. How long is the worst streak a player will meet in one session, across every window and not just one, and how often will it come?
2. Does the roll happen before the player's decision or after it, and what does each failure cost in time?
3. If the randomness has memory, how much of the bag does a player get through, and who can read what is left in it?
4. If you made this reward certain, would it still be a reward?
5. Which of your players came for the variance you are about to remove?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V4 at 0:14.
2. Candlesan, commentary, V4 at 0:27.
3. Candlesan, commentary, V4 at 0:35.
4. Candlesan, commentary, V4 at 0:40.

5. Candlesan, commentary, V4 at 0:59.
6. Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.
7. Candlesan, commentary, V4 at 1:03.
8. Vincent Diamante. Report of Rob Pardo's talk on the Blizzard approach to multiplayer design, Game Developers Conference. Game Developer, 2008, player psychology paragraph.
9. Candlesan, commentary, V4 at 2:30.
10. Candlesan, commentary, V4 at 1:17 to 1:59.
11. Candlesan, commentary, V1 at 2:05.
12. Nethaera. Real Proc Per Minute (RPPM) Trinket Changes. World of Warcraft official forums, `us.battle.net` (read from Wayback copy), 2013, section "Longer explanation".
13. Nethaera, sections "Longer explanation" and "Painfully detailed explanation for theorycrafting".
14. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "October 2003: Free Pizza and Other Hardships".
15. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
16. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., Archaeology paragraph.
17. Blizzard Entertainment. Ignite Your Summer with the Midsummer Fire Festival! `news.blizzard.com`, 2024, section "Updated Loot System".
18. Blizzard Entertainment. [UPDATED AUGUST 2] Shadowlands Season 4 Content Update Notes. `news.blizzard.com`, 2022, section "Fated Raid Rewards".
19. Ornyx. Legendary Drop Rates. World of Warcraft Forums, `us.battle.net` (read from Wayback copy), 2017.
20. Blizzard Entertainment. Dev Watercooler: Building Healthy Gameplay. `worldofwarcraft.blizzard.com`, n.d., section "Retuning Healing Spells".
21. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 13, part A, at 14:01 and 16:33; episode 21, part B, at 14:06 and 16:29.
22. Candlesan, commentary, V4 at 2:08.
23. Mark Rosewater. Mana Action. Making Magic, `magic.wizards.com`, 2011, sections "#6", "#7" and "Mana La Mancha".
24. Mark Rosewater. Nuts & Bolts 15: Structural Support. Making Magic, `magic.wizards.com`, 2023, sections "Deck Smoothing", "Mana Sinks" and "Mana Curves".
25. Helene Bergeot. Changes Starting with Pro Tour Magic Origins. News, `magic.wizards.com` (live URL 404; read from Wayback copy 2022-10-22), 2015, section "Mulligan Rule Update".
26. Ian Duke. The London Mulligan. Announcements, `magic.wizards.com`, 2019, sections "Why Change?" and "Testing the Mulligan".
27. Godot. Best-of-One Starting Hand and Mulligan Rules. MTG Arena official forum, `forums.mtgarena.com` (read from Wayback copy 2021-09-17), n.d.
28. MTG Arena Admin. Feb 14 – 0.12.00.00 Patch Notes. MTG Arena official forum, `forums.mtgarena.com` (read from Wayback copy 2021-05-07), 2019.
29. Street, "Cataclysm Post Mortem", legendary paragraph.
30. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section "Lesson 15".
31. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, Section 5.

32. Distraction Makers, episode 13, part A, at 6:42 and 7:05.

Chapter 7

Information Is a Game Mechanic

Alterac Valley was built for eighty people, and that was the trouble with it. It is a battleground in *World of Warcraft* where forty players fight forty others, and in May 2004 it was too large to test internally: a full test would have pulled eighty busy employees off their schedules. It stayed out of the shipping game, along with the rest of the fighting between players.[1]

The test finally happened in October 2004, after the gold master discs had gone to be manufactured, and it ran at half size, twenty against twenty. Half the testers were artists, unfamiliar with even mid-level spells, and half were designers, who won repeatedly and without effort.[2]

The score is the least interesting line in John Staats's entry. The zone was stocked with computer-controlled monsters and guards, and almost nobody touched them, because players had only a vague idea that killing the other side's captain won the match. The teams scattered across terrain none of them knew, "without any idea of what to expect, what to do, or where to go." Some testers never found an enemy and logged off. A player who died came back in a graveyard never seen before, and because the mini-map did not yet show allies, many ran off in the wrong direction.[3]

Nothing in Alterac Valley had been hidden on purpose. Almost everything was unknown anyway, and twenty people a side spent an hour playing a smaller game than the one that had been built.

7.1 Fog Is What Nobody Decided to Hide

Every game withholds part of its state. Some of what it withholds is the game itself, the hidden information of Chapter 5. A *Magic: The Gathering* player does not know the opponent's hand, and much of the play is acting well in spite of that. The raid facing the Prophet Skeram was meant to

work out which of three identical bosses was real from one clue, the speed at which each one's health fell (Chapter 15).[4]

The rest is ***fog***: a fact the player needs, does not have, and cannot get by playing. Where my team is. What wins the match. On a screen the two look identical. The difference is what the player can do about it. The raid at Skeram could spread its damage, watch three health bars and talk. A tester in an unfamiliar graveyard could only pick a direction.

The generous instinct after a test like Alterac Valley is to show everything: mark the captain, draw every ally, print the objective across the screen. Rob Pardo called putting everything on the interface a common pitfall, and the game kept only key information on the heads-up display and reached the auction house, a whole player economy, through a character standing in a city.[5] A fact shown at all times competes with every other fact shown at all times, so showing everything hides the few things that matter this second (Chapter 21). And some designs are made of what they hide. Tell the raid which Skeram is real and the comparing, the calling out and the cost of a wrong guess are gone.

Players do the sorting whether or not the designer does. Within a week of the public beta in March 2004, players were writing interface modifications, and one found the quests the members of a party had in common, sparing everyone the question of what to do next. It was among the first ideas the team folded into the default interface.[6] Design commentators offer a test for that sorting. If players can keep a reference page open beside the game and read the facts there, the game might as well show them. Rewarding knowledge that can only be memorized favors whoever has played longest and shuts new players out.[7] The quest mod was that reference page, written by a player, and the team passed the test by moving its contents into the game.

For every fact about the state of the game, the designer decides who knows it, when, and at what cost. Hidden information is play when the player can earn it or infer it, and fog when the player can only guess.

Design equation (a conceptual framework, not a law)

information setting of a fact = (who can see it, when it is revealed, what it costs to learn sooner)
a hidden fact the player can earn or infer is play; a hidden fact the player can only guess is fog

Three loose words of design talk become precise under it. A ***telegraph*** is a reveal the designer schedules. A tell (Chapter 15) is a reveal nobody scheduled. A ***signal*** is a reveal a player chooses to make, a visible act that says something about a private state. Alterac Valley shows the last point: a fact with no decision behind it still gets a setting.

7.2 Put the Fact Where the Eyes Already Are

A city was the purest fog in the record. In the internal alpha of September 2002 everyone who went into Stormwind got lost, because the mini-map did not work indoors. By November work on cities had stopped until navigation features existed and could be tested.[8] A missing fact had stopped the building of content.

The first answer was architectural. Stormwind Keep stood at the center so that players could steer by it. Testing showed that players always looked down, and the job moved to a cathedral spire, "in case anyone ever bothered to look up."[9] A beacon is a reveal priced at a raised camera, and players did not pay. The answers that held put the fact where the eyes already were. The indoor mini-map was tested on the art director, Bill Petras, known for poor navigation, and approved when he found his way through one of the most disorienting dungeons in the game. Maps went in during March 2004, and showing the layout of the cities dramatically improved navigation.[10] A fact only the best navigator can recover is fog for everyone else, so the test belonged with the person least able to recover it.

Players also reveal things to one another. Silhouettes were a major consideration in character design, so that a player whose eye was already overloaded with text and spell effects could tell friendly and enemy races apart by shape.[11] With art time limited, the art team put its effort into recognizable armor sets from specific places, earned by raiders and other dedicated players, and gave up wider customization.[12] The reason was production. The effect was information: armor from one place tells every passerby where its wearer has been, and the message is believable because it cannot be had more cheaply.

Then the telegraph. The game's lead encounter designers aim for the right amount of time and the right visual and audio feedback to respond to a mechanic, and where players cannot respond they turn to add-ons.[13] A telegraph can also be spent on drama, as Gruul's Shatter is (Chapter 10): a short delay in which the petrified raid can only see who stands too close, a reveal placed after the last moment anyone could use it.[14] A fact priced

in attention can also be repriced at zero by a third party. When add-ons began solving mechanics, Blizzard restricted the combat data they may read (Chapter 15),[15] along a line Pardo had drawn in 2008: mods that play the game for you are bad.[16]

7.3 Magic Hides Contents and Publishes Counts

World of Warcraft mostly fights fog. *Magic* is built the other way round. Its rules sort every card into zones and declare each one public or hidden. The battlefield and the graveyards are public. A player's hand and library, the deck being drawn from, are hidden, and nobody may look at or reorder a library.[17][18] The past is open, the present half open, the future closed to both.

The layout is deliberate in its details. Any player may count the cards left in any library, and the cards in an opponent's hand, without seeing them.[19] How much an opponent holds is a public fact about a private one. And a hand stays a hidden zone even when every card in it has been revealed:[20] a secret is defined by who has the right to look, not by who happens to know.

A hidden hand does its work through timing. Instants can be cast mid-combat or on the opponent's turn, sorceries cannot. Giant Growth is an instant because it is a combat trick, and the designers want a player to be able to bluff having it, which makes combat more dynamic. Overrun, which makes all of a player's creatures bigger at once, is a sorcery so that the opponent knows about it before choosing blockers. Sprung mid-combat it would mostly end the game.[21] The small secret gives the defender a decision. A surprise the size of Overrun leaves nothing to decide.

Untapped lands carry the next step. Spells are paid for with mana, mana comes from cards on the battlefield, and the battlefield is public. A player who ends a turn with mana unspent has shown the opponent that an instant is affordable. Whether one is in the hand stays private. The unspent mana is a signal, priced at whatever that mana would have bought on the player's own turn, and the bluff the designers want possible is that signal bought with nothing behind it. Armor cannot lie. Mana can. Thomas Schelling put the credibility of a threat at the center of his study of conflict,[22] and a signal that can be faked at the price of the real one is credible exactly as far as the real one is likely.

The opponent pays for a possible card whether or not it exists, and *Innistrad*'s design file records the payment growing too large. The mechanic that became morbid, called deathwatch in design, included spells that were

better if a creature had died that turn. Too many of them warped the environment and made players afraid to block, so they were cut back, some moved to higher rarities.[23] Every block risked switching on a spell the defender could not see. Rarity sets how believable a hidden threat is. Suppose an opponent holds five cards from a deck of forty. If two cards in the deck punish a block, the chance that at least one is in that hand is 23.7 percent, a little under one in four. If six do, it is 57.7 percent, more often present than absent. Nothing on the table differs between the two decks.

7.4 A Secret Needs Edges

Morph is the case in which a secret had to be given a shape. Richard Garfield's early card Illusionary Mask let a player put a creature onto the table face down. The trouble was unknown information about a card in play. A spell that cannot target black creatures, aimed at a face-down creature that is secretly black: the opponent cannot know whether the spell is legal, and players fell into loops of questions the owner could not answer without giving the secret away.[24]

The Rules Team's fix was the real innovation: give the face-down creature defined qualities of its own. Mark Rosewater made it a 2/2 for three mana, weak but not useless, so that players would attack with it, and varied the creatures underneath so that blocking one carried an unknown quantity.[25] The rules still read that way. A morph card may be cast face down for three mana as a 2/2 with no name and no text. Its controller may look at it, turns it up by paying its morph cost, and must reveal it when it leaves the battlefield or the game ends.[26]

Every setting is visible there. One player knows the fact. The other knows a public shell exact enough for every other rule to act on without a question. The reveal has a stated price. And the secret is audited at the end, the answer to a colleague's worry that the mechanic was open to cheating.[27] Morph hides a card by first deciding what is not hidden.

The hidden zones rest on a physical fact, which *Innistrad* exposed by breaking it. Every card has the same back, which lets one sit unidentified in a hand or a library. A double-faced card has two fronts. The shipped answer was opaque sleeves or a normal-backed checklist card standing in for it,[28] used wherever the identity must stay secret, with marks that must not show through the back, and for all of a deck's double-faced cards or none.[29] Keeping one fact hidden cost a slot in the booster and a procedure in every deck.

A fact can be hidden with nothing concealed at all. Lenticular cards keep their full text in plain view and their depth visible only to experienced players (Chapter 21).[30] The designer is also a sender who is not always heard. Players missed the link between a mechanic in *Urza's Destiny* and the rest of its block.[31] *Mercadian Masques* grouped its mechanics by creature type with no keyword, and players said the set had none.[32] A fact the designer believes is public and players cannot find is fog with a card frame around it.

Learning a fact earlier is worth something only while a decision can still change (Chapter 5), which is why *Magic*'s internal league moved from six months ahead of the public game to a year (Chapter 25).[33]

7.5 The Same Secret Is Play at One Table and Fog at Another

The settings depend on who is in the room. In a dungeon encounter built around a hidden code, groups of strangers will simply be shown the code, and a hard mode keeps the version that needs coordination.[34][35] A secret friends pass around in seconds is a wall to strangers. Private auras, introduced to make raids more interactive, show the same risk from the other side: the leads called the feature an experiment that could be worse than the problem it was meant to solve (Chapter 15).[36]

A signal also needs a receiver. A bluff cannot be written into the rules. It emerges when a decision matters, one side knows more, and the false claim is plausible. Against an opponent who does not know what could be in the hand there is nothing to bluff, and the newcomer to *Magic* who spends all the mana every turn and passes is that opponent.[37] Unspent mana says something only to a player who knows what an instant costs, so the same hidden hand is play at one table and no mechanic at all at another.

And every setting costs something. Morph needed a shell, a price and an audit, and the game's maps took two artists weeks of aligning mountains and buildings.[38] A designer can run out of rules, components and labor before running out of facts to sort.

The Designer's Test

1. For this fact about the state of the game: who knows it, when is it revealed, and what does it cost to learn sooner?

2. Can the player earn or infer the hidden fact by playing, or only guess at it?
3. If the fact is hidden, what exactly is public about it (its count, its size, its price), and is that enough for every other rule to work without a question being asked?
4. How many copies of the hidden threat exist, and at what number does the other player stop acting?
5. Who is playing together when this fact matters, and do they have a channel for passing it on?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "May 2004: The Care Bear Game".
2. Staats, *The WoW Diary*, section "October 2004: World's End".
3. Staats, *The WoW Diary*, section "October 2004: World's End".
4. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V3 at 1:34.
5. Mark Wallace. Report of Rob Pardo's keynote on World of Warcraft at the Austin Game Conference. Game Developer, 2006, paragraphs 3 and 4.
6. Staats, *The WoW Diary*, section "March 2004: Public Beta 1.0".
7. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 21, part A, at 10:50, 12:41 and 13:03.
8. Staats, *The WoW Diary*, sections "September 2002: Internal Alpha 1.0" and "November 2002: Internal Alpha 2.0".
9. Staats, *The WoW Diary*, section "January 2002: The Stitches of a Seamless World".
10. Staats, *The WoW Diary*, sections "March 2001: My First Six Months", "June 2003: A Crunchier Crunch" and "March 2004: Public Beta 1.0".
11. Staats, *The WoW Diary*, section "Character Design".
12. Wallace, report of Pardo's Austin keynote, paragraph 11.
13. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "It seems that some mechanics don't give players enough time".
14. Candlesan, commentary, V2 at 2:18.
15. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section "ADDONS & THE TRICKLE DOWN EFFECT".
16. Vincent Diamante. Report of Rob Pardo's talk on the Blizzard approach to multiplayer design, Game Developers Conference. Game Developer, 2008, section "Dealing With The Player".
17. Wizards of the Coast. Magic: The Gathering Comprehensive Rules (effective January 16, 2026). `media.wizards.com`, 2026, rules 400.2, 401.2 and 404.2.

18. Rule numbers are those of the Comprehensive Rules effective January 16, 2026. The rules change with new sets, and the numbering can move.
19. Wizards of the Coast, rules 401.3 and 402.3.
20. Wizards of the Coast, rule 400.2.
21. Mark Rosewater. Big-Picture Questions, Part 2. Making Magic, `magic.wizards.com`, 2026, answer to "What differentiates the effects you put on sorceries versus instants?".
22. Thomas C. Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.
23. Mark Rosewater. Innistrad Design Handoff Document. Making Magic, `magic.wizards.com`, 2021, section "Deathwatch comes in two versions".
24. Mark Rosewater. Wait, There's Morph. Making Magic, `magic.wizards.com`, 2002, sections "Mask Wizards" and "Magic Rules".
25. Rosewater, sections "Magic Rules" and "This Is Only a Playtest".
26. Wizards of the Coast, rules 702.37a, 702.37e, 708.5 and 708.9.
27. Rosewater, section "Magic Rules".
28. Mark Rosewater. Every Two Sides Has a Story. *magic.wizards.com (read from Wayback copy)*, 2011, sections "Double-Face the Music" and "Double-Face Off".
29. Wizards of the Coast. Double-Faced Card Rules. *magic.wizards.com*, 2011, section "Putting Double-Faced Cards in Your Deck".
30. Mark Rosewater. Lenticular Design. Making Magic, `magic.wizards.com`, 2014, section "It Depends on How You Look at It".
31. Mark Rosewater. Lessons Learned, Part 1. Making Magic, `magic.wizards.com`, 2023, section "Urza's Destiny".
32. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section "Lesson 14".
33. Mark Rosewater. New New World Order. *Making Magic, magic.wizards.com*, 2013, opening paragraph.
34. De Sousa and Sanders, interview, answer to "We haven't really explored the communication or information asymmetry mechanics".
35. The setting he names is Mythic+ dungeons run by pick-up groups, meaning parties of players who have not met before.
36. De Sousa and Sanders, interview, answer beginning "That's awesome. I know you guys have been responding".
37. Distraction Makers, episode 22, part B, at 1:09, 1:23 and 6:56 to 8:25.
38. Staats, *The WoW Diary*, section "March 2004: Public Beta 1.0".

Chapter 8

Execution Is Different from Knowledge

A mage begins to cast Polymorph, the spell that turns an enemy into a sheep. The player on the other side knows the answer, and it is one button: Pummel, an interrupt, pressed inside the cast. An interrupt is the plainest execution mechanic in *World of Warcraft*. Nothing about it needs working out. It needs to be pressed in time, which is why it sits off the global cooldown (Chapter 9).

The button is pressed in time. The Pummel lands, the mage takes its damage, and the Polymorph goes off anyway.

Nobody erred. The original servers handled players' actions in batches: a single loop processed every message once per server tick, and two actions arriving in the same batch were both treated as valid from its start. A Pummel and a Polymorph that reached the server in the same batch both happened.[1] The player knew what to do and did it, and a clock the player could not see decided otherwise. At a *Magic* table the question cannot arise, because nothing there is simultaneous. In a networked game the server's clock is one of the performers.

The sheep is a small case of a large distinction. A player can fail by not knowing what to do. A player can know and fail to do it. And a player can know, do it, and be told otherwise by the channel between the hands and the game. A designer who wants a game hard on the hands has to keep the second kind of failure and remove the third, and the first move is to stop treating the three as one thing called difficulty.

8.1 Knowing the Answer Is Not Performing It

Gruul's Ground Slam, which Chapter 10 takes apart, is the cleanest case of an answer that is fully public. The boss knocks every raider a random distance in a random direction, his Shatter then hurts everyone near you,

more the closer they stand, and the instruction that follows can be handed to every raider before the first pull: do not stand near anyone.[2] The feeling sought is musical chairs when the music stops.[3] Children know that rule perfectly, and somebody is left standing anyway. Nothing in those seconds is a puzzle.

The lava hounds at the first boss of the Molten Core (Chapter 5) make the same point from the other end. The developer running the first raid test explained the answer to the room in one line, that all the hounds had to die within seconds of one another, and forty developers still lost half the raid on every pull.[4] They had the answer. What they lacked was the ability to bring several health bars down inside the same few seconds, together.

The people who run encounter design for the game today draw the line themselves. Asked about a recent drop in dungeon complexity and difficulty, Taylor Sanders named three sources of difficulty: control of the character, such as dodging; the time and gear a player has invested; and strategy and collaboration. Not all difficulty, he said, is created the same, and he conceded that in recent expansions "we've probably trended a little bit too far into the Dodge mechanic."[5] Sanders also contrasts bosses whose mechanics are assigned at random with bosses a team can plan in full before the pull: "I think a great raid has a mix of both."[6] Easing off twitch difficulty, the leads keep mechanics of the kind raiders call the Heigan Dance,[7] a pattern learned in advance and performed in step. Inside execution, then, there are two things. Dodging reacts to what could not be predicted. A dance performs what everyone has learned.

Knowledge divides as well. A fact a player must memorize is not a decision a player must weigh. Memorized facts take a long time to learn, the time is mistaken for depth, and once they are learned the game collapses into carrying out what is known. The test is to take the memorization away and see whether anything compelling remains, and memory has one fair use, a layout redrawn every game so that nobody starts ahead.[8] The Anubisath Sentinels pass that test (Chapter 12): the pool of abilities can be studied, and each Sentinel's pick is made at the pull. Knowledge is paid in two currencies, recall and judgment.

Knowing the answer and performing it are separate difficulties, tuned separately, and performance is a fair test only after the designer has removed the failures the player cannot control.

8.2 A Shorter Window Turns Two Dials

The obvious way to make a known fight harder is to shorten the window, the time between the moment a mechanic announces itself and the moment it lands. It is a number, cheap to change, easy to compare across versions of a fight, and it needs no new art and no new rule. A team that has learned the strategy has earned a harder exam, and a faster exam on the same material looks like the obvious one to set.

It raises only the first of Sanders's three sources. Strategy gains nothing, because the answer has not changed, and gear stays where it was. A raid can become harder for a long time along that one axis while it becomes less interesting, since each added increment asks for the same thing as the last, faster. That is what the dodging admission describes.

The second failure is worse. A shorter window does not shrink the delay between screens. Gruul's first version had each client's position lag behind what the others saw, so that two raiders could both move correctly on the picture each had and both be hurt, by about half a second.[9] Half a second is a tenth of the five seconds the finished fight allows for the scramble. Cut the scramble to two seconds and it becomes a quarter. The tighter the window, the larger the share of the outcome decided by something no player controls, until the mechanic stops rewarding the faster raider and rewards the better connection. A designer turning the window turns up the demand on hands and eyes and the weight of every flaw in the channel together: the network, the warning that fires too quietly, the effect hidden under another effect. The kill rate cannot say which one is failing.

Players do not simply fail a window they cannot meet. They get help. When a mechanic gives too little time or too little visual and audio feedback, players turn to WeakAuras and other add-ons,[1011] and the add-ons came to solve the strategy side of mechanics as well, leaving the remaining challenge on dexterity and reflex, because a tool cannot steer the character out of the fire.[12] The loop closes on itself. A short window raises difficulty, players answer with tools that also take over the thinking, and the only difficulty left to add is the kind the tool cannot perform, so the next fight has shorter windows. A game whose players have good tools drifts toward testing hands, whether or not its designers meant it to. Chapter 15 follows the contest between designers and tools.

Gruul went the other way. After the Ground Slam, everybody's movement speed is slowed by 20 percent per second for five seconds, which gives the game clients more time to synchronize positions, so that everybody can play the spread reliably (Chapter 10).[13] The test became longer and slower

and measured the raid better. It did not make spreading easier in any sense a raider would recognize, since the boss still throws the raid and the gaps still have to be found. It removed one source of failure, the stale position, and left the others standing.

Who else is in the room sets the price of a miss. In converting raid bosses for groups of five, tank-switching mechanics went out, because most such groups have one tank, and healer-disrupting mechanics had to be handled far more carefully, because no second healer picks up the slack (Chapter 13).[14] In a large raid a late reaction is absorbed by somebody standing nearby. In a group of five it can end the attempt. A reaction window is two numbers, the time allowed and the number of people who can cover for whoever misses it.

8.3 The Channel Is One of the Performers

The network limits what can be asked at all. In late 2001 an animator spent a few days on procedural ocean textures, and teammates who gathered to see the result liked it. Some wanted true three-dimensional waves. The lag between client and server "would confuse the engine," because with a moving surface it could not be settled whether a creature should be above the water or below it, and tracking whether a player stood in the trough or on the crest would cost too much. There were too many engine issues, and procedural water never happened.[15]

The waves, Gruul's scramble and the double sheep are one problem. Each is a rule whose outcome depends on client and server agreeing about something while it changes quickly: a position, a surface, the order of two presses. For the sea there was no answer and the feature was cut. For Gruul there was one, which was to make the positions change slowly. For the interrupt the answer was a server that sorted messages differently, and it came years later.

The modern game runs several loops by priority and gives spell casts a high one, which made batching much less noticeable.[16] In the terms of the principle, the third kind of failure was engineered out. Execution in a networked game is never purely the player's, and the designer's job is to keep the share that is not the player's small enough that the test still measures the player.

8.4 Magic Counts Execution as Overhead

Magic: The Gathering is the control case, because it has almost none of the above. Players take turns across a table, nothing rewards a fast hand, and both players see the same board. Its clock sits at the edge of the game. Tournament rules recommend 50 minutes for a round, with a required minimum of 40, and when time is called the players finish the current turn and play five more.[17] Fifty minutes is six hundred times the five seconds of Gruul's slow. Players must take their turns in a timely fashion whatever the complexity of the situation, and a judge can be asked to watch for slow play.[18] Time presses on a tournament player's thinking in aggregate, and nothing meters a single decision or asks for a card inside a second.

The execution load *Magic*'s designers do worry about is carried in the head. Casual players at an in-house event for *Morningtide* quit after one or two rounds, though the card text had been made simpler, because cards whose effects depended on other cards in play multiplied what a player had to weigh (Chapter 21).[19] Those players could read every card, so they had the knowledge. They could not hold the board, which is a failure to perform. Bookkeeping is the same load in a plainer form: of three versions of *Kaladesh*'s energy mechanic, the one that kept counters on individual cards lost because it took more bookkeeping and did not play better (Chapter 23).[20] Tracking counted as a cost that bought no gameplay, where a *World of Warcraft* designer might count a hard dodge as the gameplay.

Herbert Simon's chooser with limited capacity is the common ground.[21] Both games spend from that capacity. *World of Warcraft* spends it on movement and timing, and its leads now talk about the mix. *Magic* spends it on tracking and interaction, and the *Morningtide* tables were an overdraft. One game sells execution and the other pays for it. *Magic* keeps nearly all its uncertainty in the shuffled deck, the hidden hand and the analysis of a board, and treats what a player must merely carry out as overhead. *World of Warcraft* makes carrying out the product, and inherits every problem of the channel through which carrying out is measured. Performative uncertainty, in Greg Costikyan's catalog,[22] is a legitimate thing to build a game from. Musical chairs has no other kind.

8.5 When the Noise Is the Game

The principle does not say to remove execution difficulty. A good raid has every type of difficulty, the leads say.[23] The hard heroic content of *Cataclysm*

suffered from a lack of alternatives, not from the difficulty (Chapter 14).[24] A game gets into trouble when the hands are the only way it is hard, or the only door in. And the two budgets, separate on paper, are paid from one head: Loatheb's healers knew the plan and abandoned it when a health bar fell (the introduction).

The limit that matters is the one that cuts against the safest half of the principle. When Blizzard rebuilt the original game as *WoW Classic*, it moved spell casts back to a low-priority loop and restored batching. Combat and the balance of fights between players had been tuned in a game that resolved spells less often, and authenticity required the old timing, with its oddities: two mages turning each other into sheep at the same instant, and a closing thought that these would be fun to see again.[25] Timing no player controls was part of the game being restored. Uncontrollable variance should go when it spoils the test the designer set, as Gruul's stale positions did. When a game has been tuned on top of it for years, the noise has become part of the test, and removing it changes who wins.

The Designer's Test

1. If every player knew the full strategy, what would still make this hard, and do the player's hands and attention control it?
2. Which failures come from the channel (network, interface, camera) and not from the player?
3. What happens to this mechanic when one person fails and nobody can cover?
4. If you shortened the window by half, would players decide better or just fail more?
5. What part of this test could a tool perform for the player, and what would be left?

Notes

1. Kaivax. Spell Batching in Classic. WoW Classic Bug Report forum, `us.forums.blizzard.com`, 2019, opening post.
2. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 0:23 and 0:59.
3. Candlesan, commentary, V2 at 0:08.

4. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section “October 2004: World’s End”.
5. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead’s Exclusive 11.2 Interview. Wowhead, 2025, answer to “We saw dungeon complexity difficulty decrease last season”.
6. De Sousa and Sanders, interview, answer to “One of the things that we had between different bosses”.
7. De Sousa and Sanders, interview, answer to “We saw dungeon complexity difficulty decrease last season”.
8. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 21, part A, at 0:15, 5:36, 16:25 and 18:41.
9. Candlesan, commentary, V2 at 1:10 to 1:38.
10. De Sousa and Sanders, interview, answer to “It seems that some mechanics don’t give players enough time”.
11. Add-ons are programs players install that read the state of a fight and display or announce things. WeakAuras is one of them.
12. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section “ADDONS & THE TRICKLE DOWN EFFECT”.
13. Candlesan, commentary, V2 at 1:48 and 2:00.
14. Scott Mercer. Blizzard Interview with Lead Encounter Designer Scott Mercer. Wowhead, 2011, answer to “Were there any specific challenges in taking a 10 or 20-player raid”.
15. Staats, *The WoW Diary*, section “November 2001: Client-Server Headaches”.
16. Kaivax, opening post.
17. Wizards of the Coast. Magic: The Gathering Tournament Rules (effective February 27, 2026). `media.wizards.com`, 2026, Appendix B and section 2.4.
18. Wizards of the Coast, section 5.5.
19. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section “Board Complexity”.
20. Mark Rosewater. Kaladesh Ingredients, Part 1. Making Magic, `magic.wizards.com`, 2016, section “Energy Conservation”.
21. Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.
22. Greg Costikyan. *Uncertainty in Games*. The MIT Press (Playful Thinking series), 2015.
23. De Sousa and Sanders, interview, answer to “Do you feel like the future of raiding relies more on throughput”.
24. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., answer to “What didn’t work out as planned?”, heroic difficulty paragraph.
25. Kaivax, opening post.

Part III

Designing Encounters

Chapter 9

A Boss Is a Behavioral Machine

The Prophet Skeram is the first boss in the Temple of Ahn'Qiraj, a raid in *World of Warcraft*, and he began as a character note. The story team described him as an intellectual prankster. Candlesan, who designed the fight, gave him two abilities on that theme: mind control, and the power to split into illusory copies of himself.[1]

Mind control looks like the more dangerous of the two. The boss reaches into the raid, takes one of its members and turns that player on the rest. A split adds enemies. Mind control subtracts a friend and adds an enemy in the same instant, and the enemy is wearing the raid's own equipment.

Then the ability met its first playtest. In Candlesan's words, "testers mostly ignored the mind controlled player." The raid was in full raid gear, and its healers absorbed one ally's damage output without strain.[2] A boss had taken over a member of the raid, and the raid carried on with what it was already doing.

The repair was a single number. He put a huge damage buff on the mind-controlled player "so that CC would be the best response."[3][4] The raid could also simply kill the teammate, and he says so with a shrug. He loved that polymorph was a solid answer, because a polymorph also heals its target, so the spell that disabled the hostile player repaired that player too. And he scaled the controlled player up in size, to add some levity, to make them fearsome, and because he thought big sheep would be funny.[5]

Skeram returns twice more: the trigger for his split in Chapter 11, and the tells that gave his copies away in Chapter 15.

9.1 Damage Is Not a Question

The first mind control dealt damage and asked nothing. Healing through it required no new assignment, no call on voice chat and no change of target,

because the healers were already casting. Every other answer required somebody to stop, find the controlled player and spend an ability on a friend. One reply cost nothing extra and worked. The others cost attention and bought nothing the first had not already bought.

So ignoring the ability was the testers' best play, and they found it at once. That is the dominant strategy of Chapter 3 at the size of a single reply: keep healing, whatever else happens. A situation with a dominant reply holds no choice, only a correct action and some ways to be wrong.

The danger was also measured against the raid's capacity, and capacity belongs to the players. A hostile teammate that would force a response from an undergeared group forced none from this one. Whether an ability asks a question depends on who is being asked.

The buff did not add options. Crowd control and killing the teammate existed in the first playtest too. It ended the dominance of the reply that needed no thought. Healing through dropped from the best answer to a worse one, crowd control rose to the top, and a fight that had asked nothing now asked who in the raid would spend which spell on a friend. The giant sheep did a second job beside the comedy: a controlled player suddenly larger than everyone else is a controlled player nobody has to search for.

The unit the designer is really placing is a ***prompt***: a piece of information, paired with a decision the player must make, under a time limit and a penalty. An encounter is the sequence of its prompts. An ability and a prompt are different objects, and the tool that builds a boss will not report the difference.

An ability is a question put to the player. If no answer to it changes the fight, it is scenery, however much damage it does.

9.2 Six Questions, and What a Missing Answer Makes

The principle works as six questions asked of one ability at a time. What information does the player receive? What decision is required? How much time is available? What actions are open? What is the punishment? How does the ability interact with the others in the fight?

Each question has its own way of going missing. Without a decision, the ability is scenery, which is what the first mind control was. Without information or time, it is noise: the player learns of it from the damage. With an empty action space, it is a cutscene with a health cost. With a punishment too small for what the group can absorb, the decision collapses

back into ignoring it. And with an unexamined interaction, the ability is quietly solved or quietly broken by its neighbors (Figure 9.1).

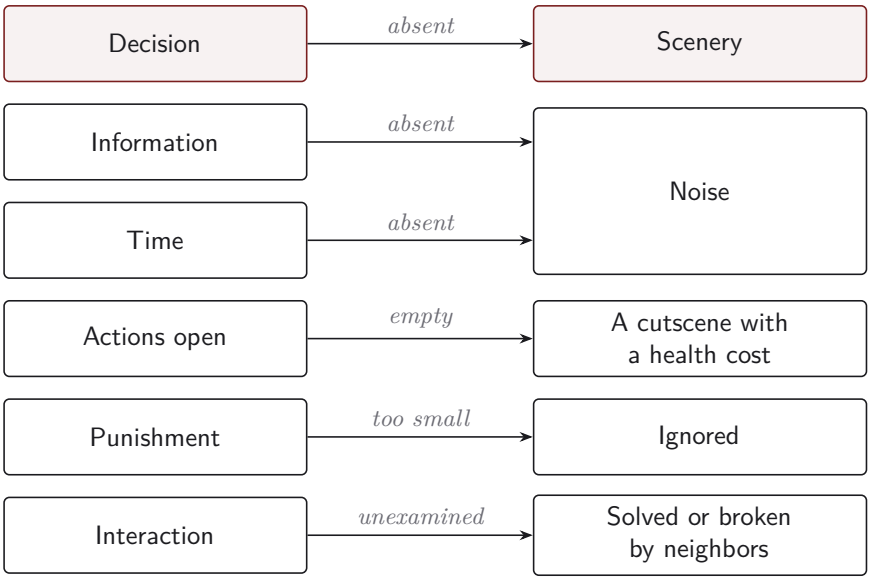


Figure 9.1. Six questions to ask of one ability, and what the ability becomes when an answer goes missing. Skeram's first mind control is the top row.

Table 9.1 runs the six questions over six abilities. The cells marked "not stated" are part of the result: designers report the decision and the punishment far more often than how long the player had.

Two columns point ahead. Shatter has a full column and a flaw: if it were the boss's only ability, the best strategy would be to give every raider a fixed spot.[6] Chapter 10 rebuilds that fight. Skeram's split fires on the boss's health and not on a timer, the choice Chapter 11 takes up.[7]

The action space is the Akil'zon pair, whose conversion to a five-player dungeon Chapter 13 tells. The old boss launched a player and interrupted them. The redesign's eagle carries a player who can still act and cast, and hands the rest of the group a target that has to die.[8] Taking away a player's buttons is the one punishment that cannot also be a prompt.

The interaction question is usually answered with a second ability. *Magic: The Gathering* faces the same overlap from the other side. Mark Rosewater's column on layering a set's mechanics starts from a constraint an encounter designer would recognize: mechanics enter the card file one at a time, so each new one has to fit the space the earlier ones left. The best second mechanic plays in a different space from the first. A later one

Table 9.1. Six questions asked of six abilities. Cells are filled from the sources cited in this chapter. "Not stated" means the source is silent.

	Mind control, first playtest	**Mind control, after the buff**	**Skeram's split**
Information	Not stated	The controlled player is scaled up in size, which also marks the target	Not stated here; Chapter 15
Decision	None observed: testers mostly ignored it	How to stop a teammate	Chapter 11
Time	Not stated	Not stated	Not stated; the trigger is health, not a clock
Actions open	Heal through it	Crowd control, polymorph included; or kill the teammate	Not stated
Punishment	One ally's damage, absorbed by healers	A huge damage buff on the controlled player	Not stated
Interaction	Not stated	Polymorph also heals its target	Chapter 15

	Akil'zon, original	**Akil'zon, redesign**	**Gruul's Shatter**
Information	Not stated	Not stated	Shatter is incoming
Decision	Not stated	The group must kill the eagle	Where to stand relative to everyone else
Time	Not stated	Until the eagle flies off; length not stated	A short delay before Shatter goes off, length not stated
Actions open	The launched player is interrupted	The carried player acts and casts normally	Spread out; on its own, a spot assigned in advance
Punishment	Potentially devastating if the tank or healer is chosen	Not stated	Damage to nearby raid members, larger the closer they are
Interaction	Not stated	Not stated	Chapter 10

can also be built to join two earlier ones, as fabricate was designed to tie together *Kaladesh*'s unconnected token and +1/+1 counter themes.[9] Shatter's partner in Chapter 10 had to ask something Shatter did not. Loatheb's spores (the introduction) were fabricate in reverse, a late piece built for the seam between the healing cap and threat.[10] Chapter 21 tells what unplanned overlap did in *Lorwyn*.

The information question has its cleanest case in *Magic* too. Early in the design of *Scars of Mirrodin*, some cards rewarded a player for having poison counters, which are normally something to avoid. Players could not tell whether attacking with their poison creatures was right, and felt they had erred when they guessed wrong, although nothing available to them indicated the correct call. The reward cards came out, and play improved because players knew to attack.[11] The cards had a decision, a punishment and all the time a turn allows, and an empty information cell. A decision with nothing to resolve it looks to the player like an error waiting to be assigned.

Last, the room. John Staats built the Molten Core, the game's first raid, in about a week, at Jeff Kaplan's request for raid content that was fast and not fancy, a quarter the size of the team's smallest dungeon. He embedded interactive runes in the floor in case the encounter scripters could use them. They went unused. A bed of hot coals near Majordomo Executus became part of that fight, and a spiral let melee characters reach Ragnaros in his lava. Beyond those, the place was wide-open space for combat.[12] Geometry can offer a prompt. Only a script can issue one, and whatever that first raid asked of forty people, it asked through what the monsters did.

9.3 The Clock Every Reply Is Counted On

No column of the table gives a length in its time row, and one fight shows what that row costs. On Reliquary of Souls, in the Black Temple, the player assigned to interrupt would often stop using abilities, for fear of being on cooldown at the instant the boss began its cast. Until *Wrath of the Lich King*, most interrupts sat on the global cooldown with everything else.[13] Ion Hazzikostas, the game's director, recalled it to testers of the *Battle for Azeroth* alpha: "That's technically a decision, but more of a nuisance than a satisfying choice."[14]

The global cooldown governed the rhythm of the game's combat. Abilities left it one at a time, until a *Legion* character could hold ten or more that ignored it.[15] Unlike the cooldowns of Chapter 4, it is shared, and it is

the unit two rows of the table are counted in: a window is so many global cooldowns long, and a reply either costs one or costs nothing.

An ability on the clock has to be chosen over the next action. An arena player at low health can spend the next global cooldown finishing an enemy or escaping. With the escape off the clock there is no decision, and the player does both.[16] Offensive cooldowns are mostly planned ahead, so freeing them bought no reaction and cost a choice: a talent became one more line in a burst macro, stacked cooldowns counted for too much in raids, and against other players a damage boost could land with the attack it amplified. The alpha put most of them back on.[17] A talent that only adds a macro line is the false choice of Chapter 2.

Interrupts stayed off, and so did the heal that saves a life: a Lay on Hands should not fail because a rotation ability was pressed half a second earlier.[18] The sorting rule is whether the ability answers something the player could not have scheduled. A later staff reply applied it to Arcane Torrent: as a silence it would stay off, but as a dispel that also generates a class resource it belongs on.[19] Testing found the rule's limit. Racial abilities and on-use trinkets went back off, because pressing several buttons in a row that only raise output is not engaging.[20]

The Reliquary interrupter was paying for what the design channel Distraction Makers calls the interaction gap, the number of actions one side may take before the other can respond. At a gap of one, every move carries a weight that can become debilitating, and players are rewarded for the smallest move available.[21] A boss cast and an interrupt form a gap of one, and the interrupter who stopped attacking was keeping the answer free.

9.4 A Zero Anywhere Gives Zero

Three ideas carry the cases. The dominant reply is the first. Affordance, defined in Chapter 4, is the second, and the encounter leads give the working version for a fight: the right amount of time and the right amount of visual and audio feedback to respond.[22] Whether the reply sits on the shared clock is part of whether there is room to make it. The third is ***interaction load***: the number of prompts a player must hold at once, and the way each overlap multiplies it. A second mechanic in a different space adds a prompt. One in the same space adds a prompt and a set of collisions.

The three compress into a line for sorting abilities. It measures nothing.

Design equation (a conceptual framework, not a law)

$$\text{prompt quality} \approx \text{what the choice changes} \times \text{whether the information suffices} \times \text{whether the time suffices}$$

The product carries one claim: a zero anywhere gives zero. The first mind control failed on the first term, since no reply changed the outcome, and it would have failed however well it was signaled. The buffed version has a first term well above zero, because crowd control, killing and healing through now lead to different fights, and a second term in a teammate grown to giant size. Its third term is blank in the designer's own account, which shows how much of a prompt goes unreported.

9.5 When the Answer Is Written Before the Pull

Designers keep fights in which most of the asking is done beforehand, and they keep them on purpose. Some fights are meant to be planned in full, players improve at them through planning and collaboration, and the current encounter leads want a great raid to mix the two kinds.[23] Under the six questions an assigned mechanic looks like scenery when it fires, because everyone already knows where to stand. The decision was real. It was made at a planning table by people who enjoyed making it. Shatter with assigned spots and a raid's weeknight strategy session are the same structure, and what separates them is whether the designer wanted the answer written down.

A prompt can also be answered by something other than the player. Add-ons solved the strategy side of mechanics, so an ability can pass all six questions on the whiteboard and still be scenery in play, read and computed by code before the player looks.[24] Chapter 15 tells that contest. The six questions describe what an ability asks, and they cannot say who will be answering.

The machine in the chapter's title has an edge of its own. A scripted opponent can read what a player does and react, but once its pattern is memorized, what remains is a dance; only another person can notice being understood.[25] The buffed mind control puts its question the same way on the hundredth pull, and by then the answer is a habit. The Anubisath Sentinels of Chapter 12 push the other way, drawing their abilities from a pool at the pull so that the raid's first task is to read what it faces.[26]

The Designer's Test

1. For each ability on your list, what does the player know at the moment it starts, and from where?
2. What is the best response, and is there a second response a reasonable player might prefer? If the best response is to ignore it, what is the ability for?
3. How long does the player have, and is that time spent deciding or traveling?
4. What can the affected player still do while it is happening, and is the reply you want on the same clock as everything else that player is doing?
5. Could the answer be written down before the pull, and do you want it to be?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V3 at 0:05.
2. Candlesan, commentary, V3 at 0:22.
3. Candlesan, commentary, V3 at 0:32.
4. CC is crowd control, the players' name for abilities that take a target out of a fight for a while without killing it. Polymorph is one of them: it turns the target into a sheep.
5. Candlesan, commentary, V3 at 0:38 and 0:42.
6. Candlesan, commentary, V2 at 0:37.
7. Candlesan, commentary, V3 at 0:58.
8. Scott Mercer. Blizzard Interview with Lead Encounter Designer Scott Mercer. Wowhead, 2011, answer to "You discussed how several of the encounter mechanics are changing".
9. Mark Rosewater. Nuts & Bolts 18: Layering Your Mechanics. Making Magic, `magic.wizards.com`, 2026, sections "Where to Begin", "The Second Mechanic" and "Necessary Elements".
10. Candlesan, commentary, V6 at 1:43.
11. Mark Rosewater. Nuts & Bolts 9: Evaluation. Making Magic, `magic.wizards.com`, 2017, section "Question #3".
12. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "October 2004: World's End".
13. The global cooldown is a short lockout shared by every ability placed on it: using any one of them starts it, and none of them can be used again until it ends. Abilities off it can be used during the lockout.
14. Ion Hazzikostas. Offensive Cooldowns Are Now On the GCD in Battle for Azeroth (alpha forum reply, as reproduced). Wowhead News (post by Anshlun), 2018, reply, paragraph 4.

15. Hazzikostas, reply, paragraph 2.
16. Hazzikostas, reply, paragraph 3.
17. Hazzikostas, reply, paragraph 5.
18. Hazzikostas, reply, paragraph 4.
19. Unsigned. Reasoning Behind Why Arcane Torrent Was Added to the GCD. Wowhead News (post by Anshlun), Blizzard forum exchange, 2018, staff reply on the silence history.
20. Unsigned. Battle for Azeroth - Racial Abilities and Trinkets Once Again Out of the GCD. Wowhead News (post by Anshlun), Blizzard forum reply, 2018.
21. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 18 at 1:04, 9:47 and 9:53.
22. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "It seems that some mechanics don't give players enough time".
23. De Sousa and Sanders, interview, answer to "One of the things that we had between different bosses".
24. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section "ADDONS & THE TRICKLE DOWN EFFECT".
25. Distraction Makers, episode 6 at 28:29 and 30:12.
26. Candlesan, commentary, V5 at 2:38.

Chapter 10

The Gruul Problem: Preventing the Obvious Solution

Gruul is a raid boss from *Burning Crusade*, the first expansion to *World of Warcraft*, and his designer's account of him begins at a children's party. "I designed this raid boss to feel like a kid's party game," Candlesan says.[1] The game was musical chairs, and the part he wanted was one instant: the music stops and everyone scrambles for a seat, a frantic mix of panic, desperation and excitement.[2]

The specification names no damage figure, no duration and no role. It names a second at a party and asks for it on demand.

The first ability he gave the boss was meant to deliver that second alone. Shatter deals damage to every other raid member near you, more the closer they stand. When Shatter is coming, everybody spreads out, fast.[3] A signal goes off, a crowd breaks apart, and each person has to find a place before the penalty lands.

He did not stop there. If Shatter were the boss's only ability, the best raid strategy would be to assign each raider a fixed spot to stand in when Shatter is incoming.[4] A raid leader would mark places on the floor, and the scramble would become a short walk to an appointment. "But what's the fun in that?" In musical chairs the children walk in a circle while they wait for the music to stop, so nobody can pick a chair in advance.[5] The walking keeps the chair from being reserved, and Shatter alone had no walking in it.

Chapter 8 used the finished fight to separate knowing from doing. This chapter rebuilds it in the order it was built, because the order is the lesson.

10.1 A Rule a Seating Chart Can Answer

The designer wants improvisation, and the rule as written can be answered with a chart drawn before the fight. The test he wanted is to find space, under pressure, among other people also looking for space. The test Shatter alone sets is to remember a mark and walk to it. The first happens during the fight and differs on every attempt. The second happens once, the night before, on a diagram, and afterward the players only carry out its result. Call a mechanic of the second kind ***precomputed***: its best response can be fully worked out before play begins. It is the single-mechanic case of what Chapter 5 called solved.

Nothing about the chart is cheating or lazy. A raid that assigned spots would play Shatter exactly as well as it can be played. The trouble is that the best play has no scramble in it.

Two patches sit within easy reach. Raise the damage, so that standing close is fatal. Shorten the warning, so that the stroll to the mark becomes a sprint. Each changes one number and sounds like the feeling being asked for: more danger, more hurry.

Both fail, for the same reason. A spread check in a room of fixed size, with a trigger the raid can see coming, has a stable answer, and a stable answer needs working out once. Raising the damage makes a mistake costlier, which makes the chart more valuable, since the chart is what prevents mistakes. Shortening the warning tightens the walk equally for players with marks and without them, so the player with a mark still wins, and the raid answers by standing nearer its marks all the time. Stakes and speed are properties of the test. The seating chart is a property of the information. A raid that knows before the pull where everyone will stand when the signal comes can solve the test at any stakes and any speed.

The party game had already shown where to look. What protects the chairs is the circle.

10.2 Scatter the Raid, Then Pay for the Scatter

So he made the raid walk the circle, by force. A second ability, Ground Slam, knocks every player a random distance in a random direction.[6] When a player lands, the mark on the floor is somewhere else, and so is everybody it was measured against. Now the raid has to improvise its spread.[7]

Then latency. *World of Warcraft* is a client and server game: each client sends its player's position up to the server, and the server sends everybody's

positions back down. Where you stand lags behind where other people see you.[8] In 2007 many players were still on dial-up, which pushed the average delay much higher.[9] Chapter 24 takes up the technical side.

Lay that over a room of people just thrown in random directions and told to get apart. On Candlesan's screen he finds a nice empty spot. Half a second later somebody walks up next to him, and on that person's screen he was the one who ran over.[10] Each player did the right thing with the picture each had, and the pictures did not match. Both get hurt, each blames the other, and neither can do anything different next time, because the mistake was in neither player's hands. His verdict on the spread as it played: "super unfun."[11]

That is a different failure from the seating chart. The chart made the test too easy to be worth taking. Stale positions meant the test no longer measured the player.

The two failures together give the principle. **A mechanic meant to test adaptation has to keep its answer from being worked out in advance, and the uncertainty that replaces the plan has to be one the player can act on.**

A player in that room faces two kinds of not knowing. Where will the knockback drop me, and who will be near me when I land? I can look, decide and move, so that kind is the game. Are the positions drawn on my screen where those players are now? Looking cannot settle that, and the next look will be no better. The first kind tests the player and the second tests the connection. The two halves pull against each other: every step away from a prepared answer asks the player to read more from the screen in less time, and so leans harder on the screen being right.

10.3 The Slow, the Stone and the Clock

The correction for latency was another ability. After Ground Slam, everybody's movement speed is slowed by 20 percent per second for 5 seconds.[12] If each second takes a fifth of the starting speed, a raider who lands running is at a standstill when the count ends. If each takes a fifth of what is left, about a third of the speed remains (0.8 to the fifth power is 0.33). Either way the raid ends the five seconds far slower than it landed.

To a player it reads as the boss tightening a grip. Its working reason concerns the machines: slowing everybody gives the game clients more time to synchronize positions, so the spread can be played reliably.[13] A slow player covers less ground in the time a position takes to reach the server

and come back, so what another screen shows is nearer the truth. The slow is a pause, dressed as a threat, in which the screens come to agree. The early seconds, when players are still quick, are for running. The late ones, when they are nearly still, are for the last adjustment, made against positions that can be trusted. As a bonus, the five seconds added drama.[14]

The fix had a shape this book keeps meeting. Ground Slam solved a problem of knowledge and made the fight depend on something the old design never needed: every client agreeing, within a fraction of a second, about where every player stood. With assigned spots, a delayed picture of a player on a mark is still correct. With a random scatter, the same delay puts people in the wrong place. The slow paid for that dependency at five seconds per cycle, and brought a consequence nobody ordered, the drama of watching your character lose speed with the gap still a few steps off.

After the slow, everybody turns to stone, and then comes one extra detail: a short delay before Shatter goes off.[15] Nobody can move. Candlesan compares it to a cartoon coyote hanging in the air past the cliff edge, a moment for players to look around in suspense, wondering whether they are in trouble and why that warlock is so close.[16] The delay repairs nothing. It finishes the party game. In musical chairs the instant after the scramble, when one child is plainly left standing, is part of the game, and the delay gives each raider that instant: all the information, and no action left.

One problem remained, and it came from the room. Since the room does not change size, a spread gets easier the more people die.[17] A raid that loses half its members to the first Shatter meets the second with half the crowd, so failing the test makes it easier for whoever is left. To keep the fight from being trivialized, he added a soft enrage: 15 percent extra damage from the boss every 30 seconds.[18][19] A raid with half its damage dealers gone takes longer to finish, and time is now what kills it. The captioned transcript breaks off mid-sentence here, with the amount described as just enough that everybody has an incentive to stay. Chapter 11 returns to this clock as a case where time is the right trigger, and Chapter 13 to the emptier room as an error that helps the group.

Table 10.1 lays the five pieces side by side. One feeling took five rules, and only the first and the fourth are about the feeling. The other three are repairs.

10.4 Three Ideas Under the Chain

Each link of the chain carries one idea (Figure 10.1).

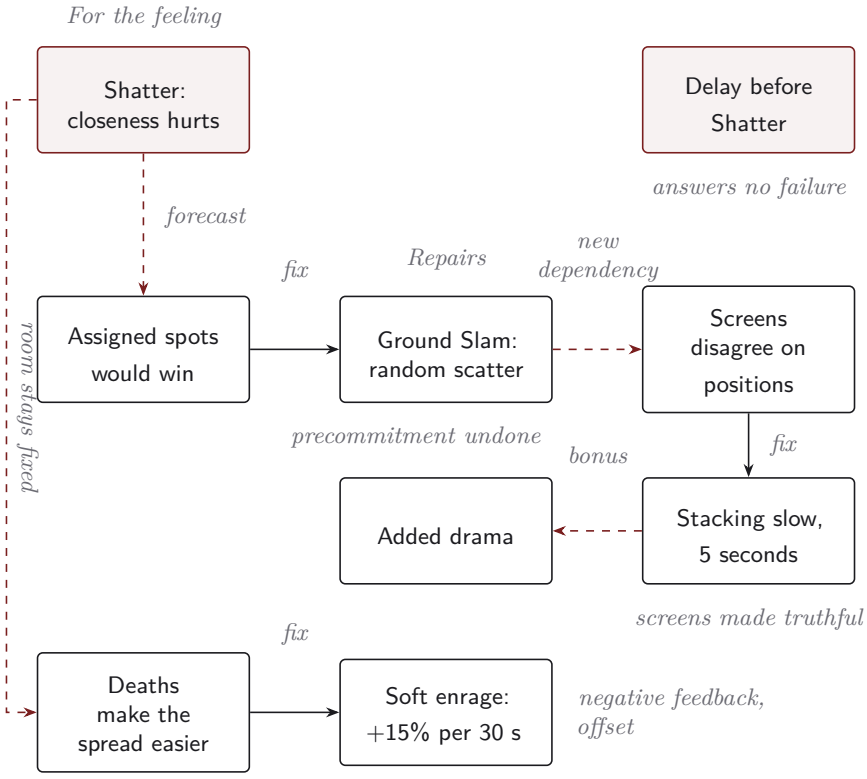


Figure 10.1. Gruul built one repair at a time, from Candlesan's account. Solid arrows are pieces the designer added; dashed arrows are what followed, whether forecast, network or players. The italic notes are the three ideas attached to the links.

Piece	What it was for	What it depended on	What it broke or left open
Shatter	The scramble: punish standing close	Players not knowing in advance where to stand	Assigned spots would be the best strategy
Ground Slam	Remove the assigned spots by scattering the raid at random	Every client showing the same positions	Latency: screens disagree and the spread is “super unfun”
The stacking slow	Give the clients time to agree	Five seconds of every cycle	Nothing broken. Added drama
The delay before Shatter	Suspense: look around, read your fate	Nothing the designer names	Nothing reported
The soft enrage	Keep deaths from making the spread trivial	A clock, 15 percent every 30 seconds	The transcript ends here

Table 10.1. The Gruul chain. Purpose and outcome follow Candlesan’s commentary (V2 at 0:20 to 2:56).

The first is ***precommitment***. A seating chart is a set of decisions made together, in advance, so that nobody has to make them under pressure, and it keeps its value only while the world it was made for is still the world when the moment comes. A designer can police the plan, with rules against standing in the same place twice, or change the world between the plan and the moment. Ground Slam does the second, which is cheaper. It forbids nothing. It makes the starting point unknowable, and the commitment loses its value on its own.

The second is performative uncertainty, from Greg Costikyan’s catalog of sources of uncertainty,[20] which Chapter 5 defined and Chapter 8 applied to this fight: doubt about whether a player can carry out, with hands and eyes, what the player already knows. Those seconds test the player only if the screen tells the truth. The slow buys a truthful screen in the test’s own currency, time.

The third is feedback. In a room of fixed size each death makes the spread easier for the survivors, so failure pushes back against the difficulty that caused it. A loop that opposes its own cause often keeps a struggling group in the fight. Here it would have dissolved the test, letting a raid

pass the party game by dying until the room was roomy. The soft enrage counters on a different axis: it leaves the space alone and charges for the time a smaller raid needs, and a clock is one thing a player's connection cannot spoil.

10.5 Poison Took Fourteen Years

Magic: The Gathering has no network to fall out of step, and it has chains of the same shape. The longest in this book is poison. Poison counters are a second way to lose: a player who collects enough of them loses whatever their life total says. Poison had failed before, R&D resisted bringing it back, and the team that brought it back listed its defects one at a time.[21]

Two defects are the Gruul problem in card form. Poison was not frightening, since totals stayed low enough to ignore until a lethal hit. And it worked mostly with creatures that are hard to block and little else, a mechanic of the kind R&D calls parasitic.[22] For most of a game, the best response to poison was settled before the game began, and it was to do nothing. A threat whose standing answer is to ignore it has been solved in advance as thoroughly as Shatter with assigned spots.

The repair was infect: what if poison worked like the older mechanic wither, with the damage itself turning into counters? Under infect anything that makes a creature hit harder makes it deliver more poison. Even a pump spell threatens a lethal amount, so the threat is constant.[23][24] The standing answer was gone, and an opponent had to judge, turn by turn, how much poison could arrive at once. Rosewater's verdict is that it took fourteen years and that he had finally got it right.[25]

Infect then needed fences of its own. A creature with infect deals no normal damage, so it could not replace wither, and the older mechanic stayed separate.[26] Infect was confined to black, green and artifacts. Proliferate, which adds to counters already in play, was held to six cards in the set, because development had found a threshold past which it became a problem.[27] Making the threat constant meant fencing it, by color, by a kept duplicate and by a count of six. Gruul's fix leaned on a network and infect's on the mechanics around it. The shape carries across, a fix, a new dependency, a correction, and the cause does not: both *Magic* players look at the same table.

10.6 When the Chart Is the Fun

Planning is a kind of fun too. For some raiders, working out the assignment is the fun, and the game's current encounter leads say a great raid mixes bosses whose mechanics are handed out at random with bosses that can be planned in full.[28] Some fights stay plannable on purpose.

A raid also learns a fight by repeating it, and structure that holds still is what makes the tenth attempt better than the first. Randomize everything and each attempt teaches less about the next. Gruul randomizes one thing, the starting position, and holds the rest still: the room, the order of abilities, the count of five seconds.

The last limit is cost. Every block of an obvious solution is a new rule, and each rule leans on something; one of Gruul's five existed only to allow for the connections of 2007. That is affordable when a correction exists. The procedural ocean waves of Chapter 8 had none, and were cut.[29] A designer who starts down this chain should expect to finish it, and Chapter 21 prices what added rules cost. If the corrections cannot be afforded, the honest choices are to accept the seating chart and make planning the pleasure, or to pick a different feeling.

The Designer's Test

1. Can the best response to this mechanic be written on a diagram before play starts?
2. If you randomize the starting condition, what does the player need to see in order to respond, and can your system show it in time?
3. What does your fix depend on that you do not control?
4. Does failing the mechanic make the next attempt at it easier for the survivors?
5. How many rules are you now using to produce one feeling, and which would you cut first?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 0:00.
2. Candlesan, commentary, V2 at 0:08.
3. Candlesan, commentary, V2 at 0:23.

4. Candlesan, commentary, V2 at 0:37.
5. Candlesan, commentary, V2 at 0:49.
6. Candlesan, commentary, V2 at 0:59.
7. Candlesan, commentary, V2 at 1:02.
8. Candlesan, commentary, V2 at 1:10.
9. Candlesan, commentary, V2 at 1:28.
10. Candlesan, commentary, V2 at 1:38.
11. Candlesan, commentary, V2 at 1:33.
12. Candlesan, commentary, V2 at 1:48.
13. Candlesan, commentary, V2 at 2:00.
14. Candlesan, commentary, V2 at 2:06.
15. Candlesan, commentary, V2 at 2:12.
16. Candlesan, commentary, V2 at 2:22.
17. Candlesan, commentary, V2 at 2:38.
18. Candlesan, commentary, V2 at 2:44.
19. A hard enrage ends a fight outright after a fixed time. A soft enrage makes the boss steadily more dangerous, so a group that takes too long is worn down instead of cut off.
20. Greg Costikyan. *Uncertainty in Games*. The MIT Press (Playful Thinking series), 2015.
21. Mark Rosewater. Something Wicked This Way Comes, Part 3. Making Magic, `magic.wizards.com`, 2010, section "#2 We Had to Make Poison Work".
22. Rosewater, "Something Wicked This Way Comes, Part 3", sections "Problem #3" and "Problem #4".
23. Rosewater, "Something Wicked This Way Comes, Part 3", sections "Problem #3" and "Problem #4".
24. A pump spell temporarily raises a creature's power, the damage it deals. Wither came from the earlier set Shadowmoor.
25. Rosewater, "Something Wicked This Way Comes, Part 3", section "#2 We Had to Make Poison Work".
26. Rosewater, "Something Wicked This Way Comes, Part 3", section "Problem #5: Poison Works Differently on Creatures than Players".
27. Rosewater, "Something Wicked This Way Comes, Part 3", sections "#2 We Had to Make Poison Work" and "#3 There Had to Be More Than Just Poison".
28. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "One of the things that we had between different bosses".
29. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "November 2001: Client-Server Headaches".

Chapter 11

Timer or Health?

The Prophet Skeram, whose mind control opened Chapter 9, has a second trick: he splits into illusory copies of himself.[1] His designer, Candlesan, had to decide when.

A boss fight is a script, and somewhere in it a condition says what fires the split. The choice as he faced it was a fixed timer or fixed percentages of the boss's health, and both, he says, are valid.[2]

Put the candidates side by side. Under the first, a clock starts with the fight and Skeram splits each time it comes around, every 60 seconds, say. The interval is invented, chosen to have something to count. Under the second, nothing happens until the raid has removed a set share of the boss's health, and then the copies appear. The copies are the same in both scripts. So are the prophet, the room, and whatever a raid has to do once several prophets are standing in it. One condition differs.

He chose health. Skeram splits at 75, 50 and 25 percent, because he wanted the illusions to "highlight technical execution regardless of the raid's DPS."[3][4]

The word carrying the weight is regardless. Damage is the number a raid's equipment raises, and he wanted a mechanic that number could not reach. A clock and a health bar look like two ways of scheduling the same event. They are two different tests.

11.1 A Clock Pays for Gear Twice

A raid's damage is not one number. It differs from raid to raid on the same night, and for any one raid it climbs, because the game pays for kills in equipment. Skeram has to work for the raid that can't quite keep up and for the raid that has overgeared the fight, arriving with better equipment than the encounter was tuned against.[5] The split is where the fight asks its question, with raids spreading damage across the prophets and calling out over voice chat which ones are losing health faster.[6] Whatever fires it has

to bring that moment to a raid on its first night in the temple and to the same raid long after it has outgrown the place.

The clock is the obvious choice. It is the simplest condition a script can state, it needs to know nothing about how the fight is going, and it brings urgency free, because the next wave comes whether or not the last one is dead. Candlesan names its real virtue: “spawning minions on a timer makes it a DPS check,” and it is very noticeable when a raid cannot keep up.[7] A raid buried under copies knows what it was short of and roughly by how much. The clock grades the raid in a currency the raid already tracks.

The property that makes a timer easy to read is the one that wears it out. With an overgeared raid, he says, you clear the minions faster and get more time on the boss, and the fight snowballs into being easier still.[8] The first effect is harmless, since better gear makes everything die faster everywhere. The second is the problem. Time not spent on copies is time spent on the boss, so the boss dies sooner, and a shorter fight holds fewer turns of the clock. Under a timer, gear is paid twice: once in how long each wave lasts, and once in how many waves there are.

A small model puts numbers on the second payment, with round invented figures rather than Skeram’s real health or any raid’s measured output: a boss with 1,000,000 health, raids dealing from 2,000 to 12,000 damage per second, and copies that arrive either every 60 seconds or at 75, 50 and 25 percent. All damage goes to the boss, so the model leaves out the time spent killing copies and understates the snowball.

At 2,000 damage per second the fight lasts 500 seconds and the timer delivers 8 waves. At 5,000 it lasts 200 seconds and delivers 3. From 9,000 upward the boss is dead in 111 seconds or less, and the raid sees a single wave. Damage rises six times across the chart and the mechanic shrinks to one eighth of its size. The threshold line does not move: 3 splits at 2,000, 3 at 12,000, 3 everywhere between.

At 5,000 damage per second both rules give 3 waves, so a designer who tested only at that level of gear would see two scripts that behave alike. They agree there and nowhere else. Below it the timer is the harsher rule, handing the weakest raid 8 waves where the thresholds hand it 3. Above it the timer is gentler, down to a third of the threshold count. A clock does not merely let strong raids off. It stretches the distance between raids in both directions, and the raid already struggling gets more of the mechanic than anyone meant it to face (Figure 11.1).

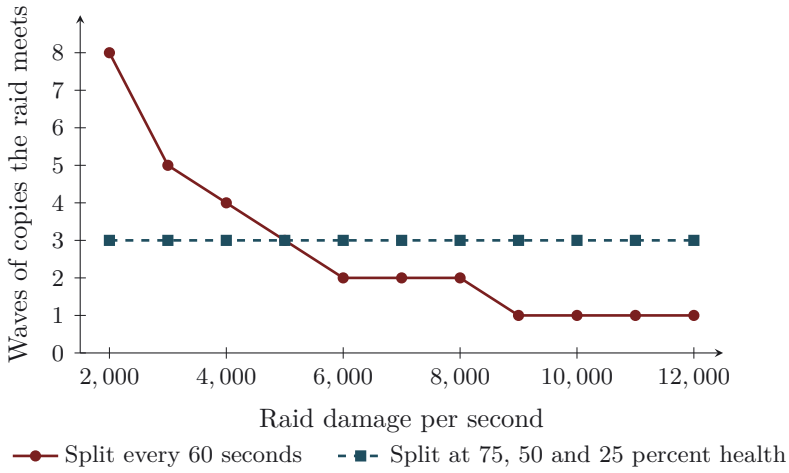


Figure 11.1. Waves of copies met on an illustrative boss with 1,000,000 health, as raid damage rises from 2,000 to 12,000 per second. The timer gives 8 waves to the weakest raid and 1 to the strongest. The health thresholds give 3 to every raid. All damage is assumed to go to the boss.

11.2 The Trigger Decides What the Fight Measures

Players answer the test that is set. Under a timer the efficient preparation is more damage: every step to the right on the chart removes waves, and a wave that never appears asks nothing of anyone. A DPS check is passed with DPS, and at the far end of the gear curve it is skipped more than it is played.

Under thresholds, damage buys a shorter fight and nothing else. Every raid meets three splits, and the trigger sits in the raid's own hands. Guess wrong about which prophet is real, push the real one to 50 percent while the two illusions from the 75 percent split are still alive, and there are five Skerams in the room and the raid wipes.[9][10] Under a clock the next split arrives whatever the raid is hitting. Under thresholds, damage on the wrong target is what summons it. The same rule would let a raid hold its damage just short of a threshold until it was ready.

The trigger decides what the fight measures.

A trigger reads a variable, and the first thing to ask is who controls that variable. Nobody in the raid controls a clock: it is an ***exogenous*** trigger, set from outside anything the players do. The players produce the boss's health

bar themselves, one hit at a time, and a trigger that reads it is ***endogenous***, a function of their own performance. The two words come from economics. The distinction is Candlesan's, in plainer language.

An exogenous event comes whether or not the players are ready, so the way to see less of it is to finish before it comes again. It rewards speed, and speed is what equipment sells. An endogenous event cannot be outrun, because the running is what fires it. The raid that deals twice the damage reaches 75 percent in half the time and finds the copies waiting there.

That matters most in a game that never stops handing out gear. Call it ***difficulty persistence***: how much of a mechanic's question is still being asked once the players' numbers have grown. A mechanic bound to the clock has a shelf life, and the gear curve sets it. A mechanic bound to state is still asking on farm night.[11] The falling line in the figure is a mechanic wearing away as its audience improves, and the flat line is the one that persists.

The designer's snowball is a ***reinforcing feedback loop***, in which a change feeds itself: more damage means fewer waves, fewer waves mean more time on the boss, more time on the boss means a shorter fight, and a shorter fight means fewer waves again. A timer wires the raid's damage into that loop. A threshold cuts the wire, since the number of waves is no longer an output of anything (Figure 11.2).

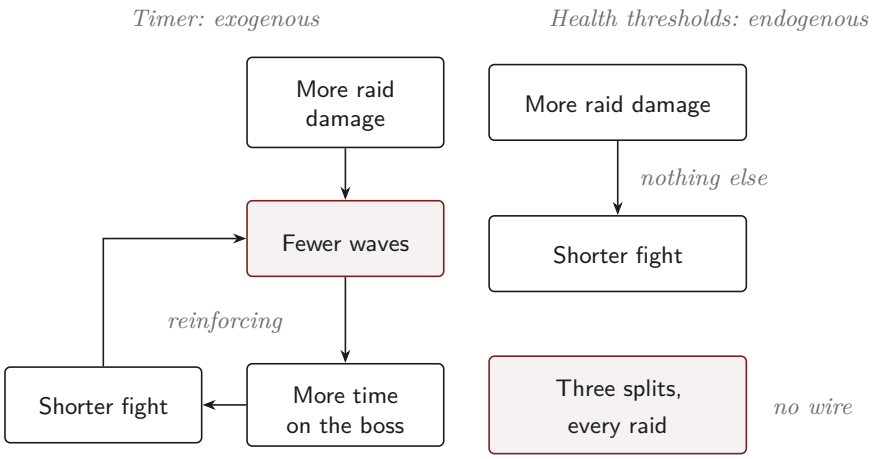


Figure 11.2. What each trigger reads. A timer wires the raid's damage into a loop that wears away the mechanic it schedules; a health threshold leaves the count at three. The loop is the systems name for what Candlesan calls a snowball.

For the simplified boss in the figure the arithmetic is exact.

Design equation (a conceptual framework, not a law)

$$\text{fight length} = \text{boss health} \div \text{raid damage per second}$$
$$\text{waves under a timer} = \text{the whole number of 60-second intervals in the fight length}$$
$$\text{waves under health thresholds} = 3$$

Damage sits in the denominator of the first line and passes straight into the second. It does not appear in the third at all. With 1,000,000 health, a raid dealing 2,000 per second fights for 500 seconds, which holds 8 full intervals. At 12,000 the fight is 83 seconds and holds 1. The third line returns 3 for both. That is the whole of the designer's regardless: the variable he wanted to exclude is absent from the formula.

A test measures whatever is cheapest to improve in order to pass it. A designer may mean a timer to test how a raid handles copies, but if more damage passes it, damage is what it measures, and a raid will supply damage long before it supplies finesse.

11.3 Clocks That Count What Gear Cannot Buy

The same designer reached for a clock on another boss, and meant it. Gruul's soft enrage, from Chapter 10, adds 15 percent to the boss's damage every 30 seconds, because in a room of fixed size a spread gets easier as raiders die.[12] On Skeram he wanted damage to have no say in how much of the mechanic a raid met. On Gruul, losing players made the mechanic easier, and damage was the one thing a depleted raid had less of. A clock that charges for time charges that raid most. The clock is the right instrument when time is the thing the designer wants to cost something.

A third kind of clock counts something gear cannot shorten. A spawn timer counts the length of the fight, and gear shortens the fight. Loatheb's cap of one heal per healer per minute (the introduction) counts what each healer may do, and no piece of equipment gives a healer a second heal inside it.[13] Timer against health is a rough label for the real question: can the players' own output move the thing the rule reads?

Both answers are still in use. Asked whether the future of raiding lies in throughput checks or mechanical ones, the game's lead encounter designers answered that a good raid has every type of difficulty, and that they lean toward strategic mechanics over twitch ones.[14] In the same interview Skeram

comes up by name, as an earlier mechanic built on players talking to one another.[15]

11.4 A Trigger Has a When as Well as a What

Magic: The Gathering has no health bar worn down by a raid, but it has triggers, and the werewolves of *Innistrad* show a second dimension of them. The goal was for both players to influence when the double-faced creatures turned. In the first design, the card checked at the end step whether no spells had been cast that turn, and if none had, the Human became a Werewolf.[16] Players learned to slip their spells in after the check. The fix moved the check to the start of the next turn.[17]

The condition never changed. It was still a turn in which nobody cast a spell. What moved was the instant at which the card looked, and with it what a player could do about the answer. A check that fires before the period it judges has ended leaves a gap behind it, and players move into the gap. The return trip had a fault in what it read, a count both players fed, which Chapter 13 takes up.

Mana is a clock the player partly controls, since a player ordinarily adds one land a turn, and the first kicker cards were small creatures that became big ones later in the game.[18] A turn of *Magic* is not gear, though. Nothing makes turns pass faster for the better-equipped player, so the decay that kills a spawn timer has no equivalent on a card. What carries over is the werewolves' point: the variable a trigger reads, and the moment it reads it, decide which behavior the trigger pays for.

11.5 Persistence Protects the Mechanic, Not the Puzzle

The principle says what a trigger measures, not which measurement a fight should take, and the chart is not a ranking. A DPS check is a legitimate test, and its failures are legible in a way thresholds cannot match. A raid that fails a timer can tell what it lacked. A raid that fails Skeram's split has more to untangle: who hit the wrong prophet, who called it wrong, whether the call was heard.

The health trigger also hands the pacing of the fight to the raid. A fight with no clock anywhere has no upper bound on its length, and a raid can take a threshold fight as slowly as its healers can sustain. A designer who

gives up the clock as a trigger often wants it back as a boundary, and a soft enrage beside a threshold phase gives it back.

The largest limit is Skeram's own. Thresholds guaranteed that every raid, at every level of gear, would stand in front of the split three times, which made the hidden-identity puzzle the one part of the fight worth breaking. Candlesan wanted the copies' halved health to be the only way to tell them apart. Testing could not disguise the real boss, and after launch players found tells the testing had missed.[19] Chapter 15 goes through them. "There's one part I never got right," he says, and that part sat inside the mechanic his trigger had protected.[20] A puzzle met three times on every kill by every raid in the game is searched more thoroughly than its designer can search it. A trigger can guarantee that players arrive at the question. What they do when they get there is a different design problem.

The Designer's Test

1. What variable does this trigger read, and who controls that variable?
2. If the players' output doubled, how many times would they meet this mechanic?
3. What is the cheapest thing a group can improve in order to pass, and is that the skill you meant to test?
4. When a group fails, can it tell from the failure what it lacked?
5. If the mechanic never goes away, what will players do to it in month six?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V3 at 0:05.
2. Candlesan, commentary, V3 at 0:58.
3. Candlesan, commentary, V3 at 1:21.
4. DPS is damage per second, the usual measure of how fast a character, or a whole raid, removes health from an enemy.
5. Candlesan, commentary, V3 at 1:10.
6. Candlesan, commentary, V3 at 1:45.
7. Candlesan, commentary, V3 at 1:05.
8. Candlesan, commentary, V3 at 1:12.
9. Candlesan, commentary, V3 at 1:51.

10. A wipe is a failed attempt: the raid dies and starts the fight again from the beginning.
11. Players say a boss is on farm when their raid kills it routinely, week after week, for its loot.
12. Candlesan, commentary, V2 at 2:38.
13. Candlesan, commentary, V6 at 1:25.
14. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "Do you feel like the future of raiding relies more on throughput".
15. De Sousa and Sanders, interview, answer to "We haven't really explored the communication or information asymmetry mechanics".
16. The end step is the last part of a player's turn. Some spells can still be cast during it.
17. Mark Rosewater. Innistrad Design Handoff Document. Making Magic, `magic.wizards.com`, 2021, section "Werewolves".
18. Mark Rosewater. Here's the Kicker. Making Magic, `magic.wizards.com`, 2007, section "Just For Kicks".
19. Candlesan, commentary, V3 at 1:34 to 2:29.
20. Candlesan, commentary, V3 at 1:31.

Chapter 12

Coordination Changes Everything

Four giant Anubisath stand at the entrance to AQ40, the larger of the two raids of *World of Warcraft*'s Ahn'Qiraj temple. Players call them the Sentinels. Their designer, who publishes as Candlesan, gave them eight abilities built in pairs, and each pair is a question. Is this one bad for melee or for casters? Should it die early or late? Does it reflect fire or frost? Should the raid stand close or far? Each Sentinel takes one ability from the pool of eight at the moment the group is pulled.[12]

So the fight opens on a puzzle that did not exist a second earlier. The raid has to read what it is up against, reposition, and pick a kill order, all in the first moments, all together.[3]

The idea was older than the raid. In 2004, at Blizzard North, the team designing monsters for *Diablo III* noticed that the game's enemy ranks lined up with its loot ranks, and that the loot had one more category, set items, which grant a bonus when worn together. They asked what a set monster would be. The answer they built was a pack of four enemies with one ability each, and each monster granted its ability to the other three when it died.[4] Kill the wrong one first and its trick is suddenly in three places. The last survivor carries everything its companions knew.

The team prototyped it. "It was not fun."[5] Then Blizzard North was shut down, *Diablo III* was put on hold, and Candlesan moved to Irvine, where his first assignment was the bosses and trash of AQ20 and AQ40.[6][7] At the entrance he found the four guards, and he recognized the problem he had left behind. "Every single reason set monsters failed in Diablo was solved right here."[8]

12.1 The Prototype Failed Because of Its Host

A set monster pack is a question dressed as a fight: in what order should these four die? The rule adds no damage and no new monster. It only rearranges the consequences of a choice. For the question to exist, three things must hold. Players must see which monster holds which ability, or there is nothing to rank. They must be able to kill one and leave the others standing, or there is no order to choose. And if several people are fighting, they must settle on the same order, or the choosing is noise.

The prototype failed all three, and none of the failures was in the rule. In *Diablo* players kill everything with area attacks as fast as they can, so four monsters fall to one blast and the order of their deaths is an accident. They do not read buff and debuff icons carefully, so the one fact the mechanic is about sits where nobody looks. And they are not on voice chat deciding whom to kill first. Who cares whether a dying enemy passed its ability on? The verdict was blunt: "Nobody even noticed."[9]

World of Warcraft supplied the three missing parts, and the designer built none of them. Its buff and debuff display is one players can read. Raiders of the period were always on Vent or TeamSpeak, two voice programs of the period. And it is a single-target game in which kill order can decide a fight.[10] Three failures, three repairs, and they map onto the three verbs of the Sentinel fight. Read is the icons. Pick a kill order is single-target combat and a voice channel. Reposition is new, the consequence of the pair about standing close or far, and it too is something a raid does together or badly. Chapter 23 draws the lesson for prototypes: keep the reasons a test failed.

A failed prototype invites one of two verdicts. The idea was bad, so close the file. Or the idea was good, so make it louder: bigger transferred abilities, harsher deaths, a tutorial message. The two disagree about the answer and agree about the question. Both treat quality as something an idea has, the way a monster has hit points. Had that been true, the set monster would have weighed the same in both games. Most ideas should die the first way, and cheaply. The set monster shows what the file has to record when they do: not the verdict, but the reasons, because the reasons may describe the game and not the idea.

12.2 Quality Belongs to the Pair

The thing that was dull in *Diablo* and worked in *World of Warcraft* was never the rule about inherited abilities. It was the rule together with its host. Candlesan's own closing line is that an idea like set monsters is neither good nor bad on its own.[11]

A mechanic's quality cannot be judged apart from the system that hosts it.

Call this ***system fit***. A prototype measures a pair, and its verdict says less than it seems to about either member. *Magic: The Gathering*'s designers apply the same test to cards: one of Mark Rosewater's questions for a set in progress is whether each card makes the set better, because nothing in a set is judged in isolation.[12] A card he likes that does not improve its set has to change. His procedure for assembling a set follows from the same idea: mechanics go into the file one at a time, and the second has to fit with the first.[13] *Magic*'s own set monster is energy, cut from two sets for fitting them badly before *Kaladesh* was built to hold it, a history Chapter 23 follows.[14]

The Sentinels make the principle specific, because the three things the new host supplied are one thing seen from three sides. A display everyone can read is shared state. Voice software is a channel. Single-target combat is a decision that pays only when the group makes it together. Shared state, a channel and a joint decision are the equipment of ***coordination***, and the set monster was a coordination mechanic from the start, waiting for players who had the equipment.

The eight abilities are built for that equipment. A pair is the natural unit of a call over voice chat: short enough to say, binary enough to act on. Melee or casters tells half the raid it is in danger. Early or late is the kill order itself. Fire or frost tells some casters to stop. Close or far tells everyone where to stand. And because the pick happens at the pull, the reading and the agreeing happen at the pull. The fight tests the equipment every time it starts. A raid whose leader is slow to call the order loses the seconds the pairs were written to fill.

12.3 Remove the Equipment and the Mechanics Break

If the equipment is what makes such mechanics work, taking it away from the same game should break them. *World of Warcraft*'s developers ran that test on themselves in October 2004, in the Molten Core attempts Chapter 5

tells. The designers described each fight in a chat window that was ignored, all communication ran through that chat, there was no voice software and no raid interface, and players saw health bars only for their own party. The volunteers were replaced by the forty best players from the QA department, a group that worked together in the same room and came much closer to how a real raid performs.[15] The content had not changed between the two groups. The QA players were also more skilled, so the comparison is not clean. It still isolates what the first group lacked: forty people typing into a window nobody read.

Two decades on, the dependence is policy. In 2025 the game's lead encounter designers were asked about mechanics that rely on communication, and Taylor Sanders named as precedent Prophet Skeram, the AQ40 boss whose raid the Sentinels guard. The designers said that puzzles built on communication and on players holding different pieces of information suit coordinated groups of friends better than pick-up groups of strangers. Drew De Sousa described a dungeon in which a pick-up group in the Mythic+ mode is shown the code at the center of such a puzzle, while the hard mode keeps the version that requires talking, and he said the team would not do the coordinated kind everywhere.[16] The mechanic is matched to the group that can host it, one difficulty setting at a time.

The designers have also made coordination the measure of difficulty. In 2026 the lead encounter designer Dylan Barker described the climb from Normal to Heroic to Mythic as harder on a communication axis and a coordination axis, not as more bullets aimed at the raid.[17]

The host does not hold still. The top teams commissioned add-ons that did the reading and assigning for them, designers began building mechanics with those tools in mind, the results trickled down to ordinary groups, and the pre-patch for *Midnight* restricts the combat data the tools can use.[18] A mechanic that asks players to read and agree is worth something different once software does both. Chapter 15 follows that contest.

12.4 A Coordination Problem Is Solved by Agreeing

Picture the raid in the seconds after the pull. Several kill orders may be acceptable, and the raid does not need the best one. What it cannot afford is half its damage on one Sentinel and half on another. Each player's payoff depends less on choosing well than on choosing what the others choose. Raiders who have each worked out the ideal order are no use to one another

if they worked out different ones, and a leader who names a mediocre order on the voice channel beats them all, because forty people acting on one plan outperform forty acting on forty better ones. This is a ***coordination problem***, the kind Thomas Schelling's work made familiar.[19]

That is why the equipment has three parts. The facts must be visible to everybody. Each player must be able to count on the others having seen them, which a shared display provides and a private hunch does not. And the choice must be announced so that everyone knows it has been made. The October 2004 test had the first in a thin form and almost none of the third: instructions typed into chat did not become shared knowledge by being available.

The display part is a matter of affordance, which Chapter 4 defined as what a thing shows a person about how it can be used. An icon above an enemy's head affords reading only if it is distinct, stable and placed where the player's eyes already are. *Diablo* had icons. Its play gave nobody a reason or a second to study them, and the difference between the two games was a system players could actually read. And the parts are complements, inputs worth more together than apart, in the way a mechanic and its host are. Energy and the set built to feed it are complements to the point where neither is worth much alone.

The three parts fit in one line. It is a framework for interrogating a coordination mechanic before building it, and it has no units.

> **Design equation (a conceptual framework, not a law)**
>
> value of a coordination mechanic $\approx$ how observable its state is $\times$ how selectively players can act on it $\times$ how easily they can agree

The form is a product because the factors do not compensate for one another. Perfect icons and a perfect voice channel, with attacks that hit everything, leave no order to choose. *Diablo* was weak on all three factors and *World of Warcraft* strong on all three. *Magic*, played as a duel, has no teammates to agree with, so only the half of the principle about hosts carries over; Two-Headed Giant, the format in which partners share one life total, is the exception, and Chapter 13 tells it.

12.5 The Principle Will Not Find the Host for You

Building difficulty out of coordination decides in advance which groups may attempt it. The code puzzle has a second version because the first shuts strangers out, and the worry is as old as the game: during the open beta of 2004 developers judged that raids looked likely to appeal to about 1 percent of the player base, one player in a hundred, an impression Chapter 5 recounts.[20] A host that supports coordination mechanics is a host only some players live in. Chapter 13 prices that inside a single group, and Chapter 14 across the population.

The sharper limit runs against the principle's own wording. "No quality apart from the system" is true as analysis and fatal as a working rule. Taken literally, it would have filed the set monster as a fact about one pair and forgotten it. Rosewater kept carrying energy from set to set because he judged it a good mechanic before any set had proved it.[21] That is a judgment of the mechanic in itself. Candlesan's judgment at the AQ40 entrance was the other kind, a judgment about the host, and it was the principle at work. But he could make it only because he still carried the idea and its three failures. A designer who had filed the first prototype and forgotten it would have had nothing to recognize when the right host appeared. The principle says where quality is found. Getting an idea to that place takes a private conviction about the idea itself, and Chapter 23 prices the keeping.

The Designer's Test

1. What must a player be able to see for this mechanic to pose its question?
2. Can players act on one thing at a time, or does your combat hit everything at once?
3. How do the players tell each other what they have decided, and what happens to a group that cannot?
4. Which part of the host system does the mechanic rely on, and who else can change that part?
5. If the prototype was not fun, was it the idea or the game around it, and what would a host that suits it look like?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V5 at 2:26.
2. To pull is to start a fight by drawing the enemies' attention. AQ20 and AQ40 are the players' abbreviations for the two Ahn'Qiraj raids; the transcript is machine-captioned and garbles the temple's full name.
3. Candlesan, commentary, V5 at 2:43.
4. Candlesan, commentary, V5 at 0:28 to 1:22.
5. Candlesan, commentary, V5 at 1:24.
6. Candlesan, commentary, V5 at 1:48.
7. Trash is the players' word for the ordinary monsters that stand between a raid and its bosses.
8. Candlesan, commentary, V5 at 2:09.
9. Candlesan, commentary, V5 at 1:30 to 1:45.
10. Candlesan, commentary, V5 at 2:15.
11. Candlesan, commentary, V5 at 2:48.
12. Mark Rosewater. Nuts & Bolts 9: Evaluation. Making Magic, `magic.wizards.com`, 2017, section "Question #2".
13. Mark Rosewater. Nuts & Bolts 18: Layering Your Mechanics. Making Magic, `magic.wizards.com`, 2026, section "The Second Mechanic".
14. Mark Rosewater. Kaladesh Ingredients, Part 1. Making Magic, `magic.wizards.com`, 2016, section "Building Up Energy".
15. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "October 2004: World's End".
16. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "We haven't really explored the communication or information asymmetry mechanics".
17. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, last Barker quote.
18. Jaydaa, report of Barker interview, section "ADDONS & THE TRICKLE DOWN EFFECT".
19. Thomas C. Schelling. *The Strategy of Conflict.* Harvard University Press, 1960.
20. Staats, *The WoW Diary*, section "November 2004: Open Beta".
21. Rosewater, "Kaladesh Ingredients, Part 1", section "Building Up Energy".

Chapter 13

When One Player's Mistake Becomes Everyone's Problem

Zul'Aman was a raid its makers were proud of and few players had seen. It had launched late in the Burning Crusade cycle, and the team felt it was one of its best.[1] So the team did something unusual with it. Zul'Aman and a second raid, Zul'Gurub, built between them for groups of ten and twenty, were reissued as Heroic dungeons for five.[2]

A raid carries several tanks and several healers. A group of five usually has one tank and no second healer. The conversion put every boss in front of a smaller and thinner crowd.

One of those bosses was Akil'zon, whom Chapter 9 met for what his mechanic leaves a player able to do. In the raid he launched a chosen player into the air, interrupting whatever that player was doing. For the dungeon the designers judged that the toss could be devastating if it chose the healer or the tank.[3]

Nothing about the toss had changed. It picked one player and took that player out of the fight for a moment, as it always had. What had changed was the number of people standing under it, and with that number, the odds that the person in the air was someone nobody else could stand in for.

The conversion produced two rules. Scott Mercer, the game's lead encounter designer, gave the first plainly: "you can't use tank switching mechanics when you know that most 5-player groups are only going to have one tank." The second was that mechanics which silence, stun or otherwise disrupt a healer had to be used much more carefully, because no other healer was available to pick up the slack.[4] A healer silenced for a few seconds among several healers is an inconvenience. The same silence on the only healer present ends the attempt.

13.1 A Mechanic Assumes Cover

The player Akil'zon threw had done nothing wrong, and the designer's question is the same whether there is a mistake or not. One member of a group stops functioning, through an error, a death or a boss's whim, and somebody decides what the other members lose. In a fight for one player the person who slips is the person who pays. In a group the bill can be sent elsewhere, and the designer chooses the address.

Two instincts set that address badly. The first treats a mechanic as portable: a toss, a silence or a tank swap tuned in a raid has a known severity, so it can move into a smaller dungeon with its numbers scaled down. The second treats harshness as a virtue: if group content is about players depending on each other, the strongest version is a rule under which any one failure ends the attempt for all. The developers who first heard such a rule, the lava hounds of their own Molten Core tests, all to be killed within seconds of one another, cheered it (Chapter 5).[5]

Both instincts place severity inside the mechanic, and the three mechanics Mercer names show why that is the wrong place. A tank swap assumes a second tank. A healer lockout assumes someone else can heal through it. The toss assumes whoever goes up can be spared for the length of the flight. In a raid those assumptions hold. In a group of five, each fails for a specific seat. Halve the silence and the lone healer is still the only healer for all of it. Severity is a product of what a mechanic does and how much cover the group has, and group size changes the second factor without touching the first.

The harsh answer fails on behavior. When the cost of other people's conduct lands on a player who has no means of covering it, the cheapest move is to leave. In the first company alpha anyone could loot any corpse, and everyone's goal became to level out of the crowded starting area and play solo, a case Chapter 1 tells.[6] A shared fate a player can do nothing about gives that player a reason to stop sharing.

13.2 Price the Slip and Keep the Stake

Group punishment is a price, and the designer sets it through redundancy.

The working idea is ***cover***: the number of players who can absorb a member's slip or do a removed member's job. With one tank, removing the tank ends the attempt. With several healers, losing one for a moment is a

nuisance. The mechanic is the same in both sentences. Its severity has to be tuned against the structure of the team that will meet it, which is why a smaller group changes bosses whose own rules did not change.

The eagle shows what a designer does with that. The redesigned Akil'zon summons eagles that pick a player up and try to fly off. The carried player can still act and cast, and the rest of the group has to kill the eagle. In Mercer's words, "the eagle mechanic still creates a sense of urgency without completely disrupting the targeted player."[7] The group-wide stake survived. Somebody is still being carried away, and everyone still has to care. What went is the helplessness, on both sides: the carried healer keeps healing, and the other four have a job that fixes the problem. One player's bad luck became a task the rest can perform.

The conversion reached past the bosses. Zul'Aman needed few changes because it already followed the team's current encounter philosophy. Zul'Gurub needed more: its ordinary creatures were redone so that five players could clear them, and some wings were closed off with doors or fallen debris.[8] Disruption was rationed, not abolished. Careful use of healer disruption is still use.

A slip reaches the rest of the group by three routes. The first is role, the conversion's route: with one occupant per role, whatever happens to the occupant happens to the function. The second is space. Gruul's Shatter, which Chapter 10 takes apart, deals damage to every other raid member near the player based on how close they stand.[9] Its twist is that a fixed room makes a spread mechanic easier as people die, so a member's death helped the group, and the designer corrected the sign with a soft enrage of 15 percent extra damage every 30 seconds.[10] The third route is synchrony. Under the lava hounds rule, the last hound sets the result however well the others were handled, and the slowest part of the group decides for all.

13.3 Threat Is a Budget the Tank Holds

A fourth thing a group can share is a budget, and it was the subject of the first raid test in the game's history. On October 15, 2004 a twenty-person test was held to see whether aggro management worked in large groups, and the aggro code was rewritten before the forty-person test that followed.[11] Threat, the game's running account of whom an enemy most wants to attack, is produced by everyone and has to be led by the tank. The introduction showed Loatheb's spores protecting that lead by letting players deal damage that generates none.[12]

Greg Street's design post on threat says what the system was for and what it turned into. Its purpose was to make fights more interesting. In practice damage dealers watched the tank's lead and held back their own abilities to avoid pulling the enemy onto themselves, and leaned on third-party meters to track the gap. The post singles out random groups assembled by the Dungeon Finder, where an undergeared tank paired with well-geared damage dealers lost the enemy again and again, a matter of numbers and not of skill, and no fun for anyone. The repair raised a tank's threat from three times damage done to five times.[13]

This is the conversion's logic applied to a quantity. The tank's lead is a cap on everyone else's output, the cap is set by one player's gear, and a stranger's shortfall becomes the group's ceiling. The designers removed the stake where a number decided it and kept it where skill does: with threat made easy, a tank who stands in the wrong place or misses a transition still wipes the party or the raid. The same post names the repair's cost. Tanks relieved of the race may be left with buttons that no longer matter.[14] The fix moved a stranger's shortfall off the damage dealers and left the tank holding fewer decisions.

Whom the group is decides how much a slip costs. The current encounter leads say that mechanics built on every member holding up one end of a puzzle suit groups of friends better than pick-up groups of strangers, and offer a pick-up version beside the hard one (Chapter 12).[15] Being wiped by a friend's slip and by a stranger's are different experiences, and the same rule is sold at two prices. Chapter 14 follows that across the population, and Chapter 17 takes up the friend and the stranger.

13.4 One Life Total for Two Players

Magic: The Gathering is usually a duel, and in a duel there is no teammate to hurt. Two-Headed Giant poses this chapter's question. Two players form a team against another pair, the team has one life total starting at 30, and the team shares no other resource.[16] Each player brings a separate deck, and life gained or lost by either teammate moves the team's single total. *Battlebond* was a set made for the format.[17]

Lands and creatures stay private, so a misplay with them costs their owner a card or a turn. Life is the one stock both heads draw on. A partner who leaves an attacker unblocked, or pays life for an effect, spends something the other was counting on. It is Shatter with two people in the room.

The format also supplies the cover. A player may block creatures attacking a teammate.[18] *Battlebond* introduced assist, which lets another player pay generic mana toward the cost of your spell.[19] These are the eagle redesign in card form: the shared stake is kept, and the partner is handed a way to act on the other head's trouble.

Even a duel can bind two people to one quantity. *Innistrad*'s werewolves first flipped back on two spells cast in a turn by any mix of players, which created odd incentives, and the trigger was narrowed to one player casting two.[20] Chapter 11 follows when the same cards check. In the same design, blue Zombies exiled creature cards from any graveyard, a broad version that created what Mark Rosewater calls weird incentives, and the redesign limited the exile to the caster's own graveyard.[21][22] Both repairs shorten an effect's reach, and reach decides who pays.

13.5 Every Player Added Is Another Chance to Slip

Three ideas from economics fit the cases. The ***externality*** is a cost paid by someone other than the person whose action created it, and Shatter is one by construction. The ***weakest link*** is a group whose result is set by its poorest member; Jack Hirshleifer's paper title sets weakest-link against best-shot provision.[23] The lone healer is a weakest link of role, and the last lava hound a weakest link of timing. The third is the shared stock each member is tempted to draw down, the tragedy of the commons since Garrett Hardin.[24] Open looting in the first alpha had that shape, and so did the tank's lead before the multiplier rose.

Weakest-link structures have an arithmetic that intuition gets wrong. Suppose each player slips rarely on a given mechanic. The chance that somebody slips grows with the size of the room, and Figure 13.1 shows how fast. The model treats each player's error as independent, which real raids are not: a confusing signal or a bad night trips several people at once.

At a 2 percent error rate per player, somebody among 5 slips 9.6 percent of the time, about one attempt in ten. Among 40 the figure is 55.4 percent, more than every second attempt. At 5 percent per player the two numbers are 22.6 and 87.1. Even at 1 percent, where each individual is nearly flawless, a room of 40 sees a slip in 33.1 percent of attempts, against 4.9 percent for a group of 5.

So group size cuts both ways. A large group has more cover in each role, which was Mercer's point, and far more chances for someone to slip, which is

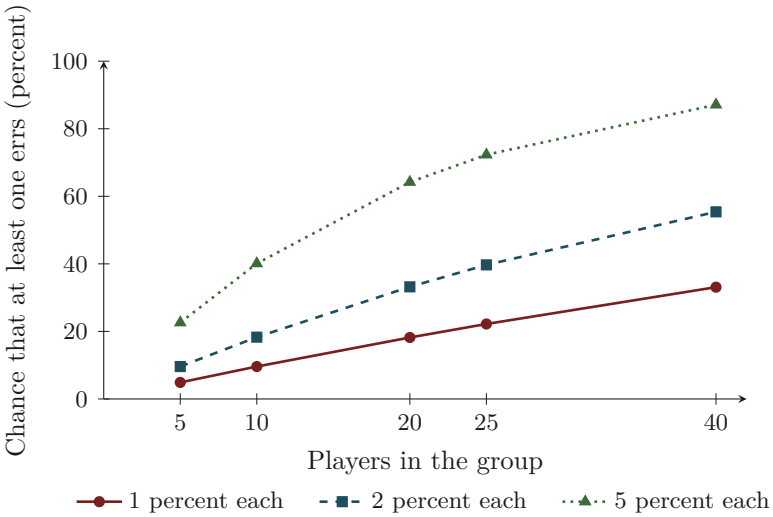


Figure 13.1. The chance that at least one of n players errs on a mechanic, for per-player error rates of 1, 2 and 5 percent. At 2 percent the chance is 9.6 percent among 5 players and 55.4 percent among 40. Errors are assumed independent.

the figure's. A rule under which any single slip ends the attempt is tolerable for five and worse than a coin flip for forty at the same 2 percent level of play. The two effects fit in one line, a framework for asking the question that returns no number.

Design equation (a conceptual framework, not a law)

$$\text{cost to the group} \approx \text{chance that some member slips} \times \text{share of the group the slip reaches} \div \text{cover}$$

Mercer's conversion lowered the first term, since five players slip less often than ten or twenty, and his two rules guarded the last. The lava hounds rule set the reach at everyone and the cover at nothing, in a room large enough to make the first term high. The developers who cheered it watched beta groups wipe the next month and judged that raids might appeal to 1 percent of the player base (Chapter 5).[25] The stakes that produced the cheer and the stakes that produced the doubt were the same stakes.

The limit runs the other way. Turn redundancy all the way up and one player's error stops mattering, and so does the player. In the first Alterac Valley test, which Chapter 7 tells, some testers never found enemies

and logged off.[26] A structure in which no one's absence costs anything has stopped asking anything of anyone. Gruul's designer added a clock to an error that helped the group, Mercer rationed disruption for five and kept it, and Street defended the wipe that follows a missed tank transition. Shared stakes are much of what group content is for. The design question is what to charge, and to whom, and whether the people charged have any way to pay.

The Designer's Test

1. When a player slips on this mechanic, or is removed by it, who takes the damage?
2. How many players in the group can cover for the one who is removed, slowed or killed?
3. If you halve the group size, which of your mechanics ends the attempt?
4. Does a death make the mechanic harder, easier or the same for the survivors?
5. What kind of group, strangers or friends, is this punishment priced for?

Notes

1. Scott Mercer. Blizzard Interview with Lead Encounter Designer Scott Mercer. Wowhead, 2011, answer to "How did Rise of the Zandalari come about?".
2. Mercer, interview, answer to "Were there any specific challenges in taking a 10 or 20-player raid".
3. Mercer, interview, answer to "You discussed how several of the encounter mechanics are changing".
4. Mercer, interview, answer to "Were there any specific challenges in taking a 10 or 20-player raid".
5. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "October 2004: World's End".
6. Staats, *The WoW Diary*, section "September 2002: Internal Alpha 1.0".
7. Mercer, interview, answer to "You discussed how several of the encounter mechanics are changing".
8. Mercer, interview, answer to "Can you give us any indication of the how the major bosses".
9. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 0:23.
10. Candlesan, commentary, V2 at 2:38 to 2:44.

11. Staats, *The WoW Diary*, section "October 2004: World's End".
12. Candlesan, commentary, V6 at 1:43.
13. Greg "Ghostcrawler" Street. Dev Watercooler – Threat Level Midnight. World of Warcraft (`worldofwarcraft.blizzard.com`), n.d., sections "Why have threat?", "Why not have threat?" and "Upcoming changes".
14. Street, "Threat Level Midnight", section "Long-term changes".
15. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer to "We haven't really explored the communication or information asymmetry mechanics".
16. Wizards of the Coast. MTG Two-Headed Giant. Magic: The Gathering (`magic.wizards.com`), 2026, section "Play Rules/Modifiers".
17. Matt Tabak. Battlebond Mechanics. Magic: The Gathering (`magic.wizards.com`), 2018, section "#squadgoals".
18. Wizards of the Coast, section "Play Rules/Modifiers".
19. Tabak, section "Assist".
20. Mark Rosewater. Innistrad Design Handoff Document. Making Magic, `magic.wizards.com`, 2021, section "Werewolves".
21. Rosewater, "Innistrad Design Handoff Document", section "Zombies".
22. A graveyard is a player's discard pile. To exile a card is to remove it from the game.
23. Jack Hirshleifer. From weakest-link to best-shot: The voluntary provision of public goods. *Public Choice*, 41:371–386, 1983.
24. Garrett Hardin. The Tragedy of the Commons. *Science*, 162 (3859):1243–1248, 1968.
25. Staats, *The WoW Diary*, section "November 2004: Open Beta".
26. Staats, *The WoW Diary*, section "October 2004: World's End".

Chapter 14

A Game for Everyone Is Tuned for Someone

When the raid queue of *World of Warcraft* went live with patch 4.3, Blizzard published a page of questions and answers about it, and the first answer named an owner. Raid Finder was meant mainly for players who did not already raid consistently: people whose schedules, playing time, friends or experience kept them out of organized raids. A second use was gearing a spare character without being bound by a raid lockout.[1]

Every rule on the page follows from that sentence. The queue assembles 25 players, a size the page defends as forgiveness, since in a group that large one player's error costs less.[2] The items sit below those of the Normal raid. Loot is rolled with priority for a player whose need matches the primary role, so a healer's item goes to a healer. A player can win loot from a given boss once a week, and there is no lockout, so a boss skipped this week is still waiting next week.[3]

Read as a set, those rules describe a person: no guild, no fixed night, no assurance that the other 24 know the fight, and no wish to be shut out for a week because of whom the queue picked last time.

The person who turned up was someone else. About two and a half years later, in the third of a series of design articles on raiding, Ion Hazzikostas, the game's lead designer, wrote that guild raiders had come to feel they needed to run Raid Finder every week, for its set bonuses and particular trinkets.[4] The repair followed. Raid Finder would get its own loot tables and its own item art, at a power between dungeon loot and Normal raid loot, with no set bonuses and none of those trinkets, and its rate of reward would nearly double.[5]

A mode built for players who could not commit to a schedule had become one more night on the schedule of the players who had. Nothing in the queue had malfunctioned. Its gear was worth having, and the guild raiders were the players best placed to notice.

14.1 One Setting Cannot Serve Two Commitments

The difficulty is as old as the genre. A game that never ends must appeal to audiences with very different levels of commitment, from casual players to those who play more than a hundred hours a week.[6] An encounter is where that spread meets a single setting. A raid boss has one script, one loot table and one weekly reset. A boss tuned for a guild that has practiced for a month is a different object for 25 strangers, and a boss tuned for the strangers is a different object for the guild.

The designers had two answers at hand. The first is to choose. Rob Pardo, in a 2006 keynote, pictured the audience as a ring with the core players at the center and said the dungeons were there to serve the core market and not the casual one.[7] The second is the middle setting, a value both groups can live with, and Pardo spoke against it: where the team can, it avoids compromise, because a compromise usually leaves both sides unhappy.[8]

Cataclysm tried the first answer at full strength. Its heroic dungeons and raids, the game's name for the hard versions, were tuned back up to roughly the difficulty of *The Burning Crusade*'s dungeons, for players who had asked for harder challenges. Greg Street's post-mortem records the split. Players who wanted that difficulty were delighted. Another group felt stuck, unable to earn the gear that raids required, the more so because Raid Finder did not yet exist. The content was not too hard, since some players plainly enjoyed it. The trouble was that there were no alternatives.[9] One setting chosen for the core sent a bill to everyone outside it.

The alternatives arrived in stages. Raid Finder came in the same expansion, with patch 4.3,[10] and the next expansion was to offer normal and heroic raids with Raid Finder beside them, and normal and heroic dungeons with a challenge mode for players who wanted to prove their skill.[11] The repair to a single setting was more settings. The warning had come earlier: during the 2004 open beta, developers watching groups fail judged that raids might appeal to about 1 percent of the player base (Chapter 5).[12]

The middle setting fails more quietly. Until 2014 the hardest raid difficulty was called Heroic and came in two sizes, 10 players and 25, meant as one difficulty. Tuning the two precisely against each other remained a major challenge: Lei Shen was easier with 10, Garrosh easier with 25.[13] The two sizes were two audiences, small rosters and large ones, and each fight leaned toward one of them. A difficulty meant to sit between them was missed boss by boss, in both directions. Chapter 13 gives the arithmetic underneath.

14.2 Split the Rule and the Problem Moves to the Seam

A rule shared by two audiences is tuned for one of them, and the other pays the difference. Splitting the rule moves the problem to whatever the pieces still share.

A ***difficulty tier*** is one version of a piece of content tuned for one audience; Raid Finder, Normal, Heroic and Mythic, in their 2014 names, are difficulty tiers of one raid. A ***seam*** is whatever two difficulty tiers still have in common. Three seams run between the tiers of a raid. The reward is the items worth something in more than one tier. The calendar is the lockout that decides which tiers a player may enter in a week. The roster is the number of players a tier is tuned for. Each seam is a way for one audience's game to reach into another's.

Players crossed every seam they could. Guild raiders ran the queue for its gear.[14] Most players who raided in more than one difficulty did so with different social groups: a guild on its regular nights, and, in the post's example, a weekend run in the flexible mode with real-life friends on another server, for items suited to a second specialization.[15] And in the older Trial of the Crusader raid a currency called Emblems made guilds feel obliged to keep clearing the Normal version after they no longer needed its loot.[16] An item, a friend, a currency: each crossing ran along something the two tiers still shared.

The 2014 restructuring treated all three seams at once. The old flexible mode became Normal. The old Normal, played at 10 or 25, became a single Heroic of flexible size. The old Heroic, also at 10 or 25, became Mythic, fixed at 20 players.[17] Table 14.1 sets them against the three seams.

The roster seam was cut first. A flexible size means a group of 13 need not cancel the night when one or two members are missing, and nobody has to sit out; the post calls the result organized raiding without the same feeling of obligation. A single fixed size at the top does the opposite job: the hardest version is tuned for one roster, and the designers stop trading 10 against 25.[18] Below the top the size bends to the group. At the top the group bends to the size, because that is the tier where precise tuning matters most.

The reward seam was the one that had leaked. Raid Finder has none of the pacing of progression. A guild's raid slows at the later bosses and speeds up once a fight is mastered, and that rhythm keeps organized raiding engaging from week to week; in the queue, the ninth clear of the Underhold

Table 14.1. The four difficulty tiers of the 2014 design, as Hazzikostas's post describes them. "Not stated" means the post is silent. Two Raid Finder cells come from the earlier question-and-answer page and describe the mode at patch 4.3, since the post does not restate them: the 25-player roster, and the calendar (no lockout, loot from a given boss once a week). The reward cells compare difficulty tiers only where the post does.

Difficulty tier	**Roster**	**Calendar**	**Reward**	**Stated audience**
Raid Finder	Not stated (25 at patch 4.3)	No lockout; loot from a given boss once a week (at patch 4.3)	Own tables and art, apart from the other three difficulty tiers; between dungeon and Normal power; no set bonuses or certain trinkets; rate nearly doubled	Players who cannot or will not commit to a fixed group; players who want the story
Normal (was Flexible)	Flexible; 13 given as an example	Its own weekly lockout	Better rewards than Raid Finder; not compared with Heroic or Mythic	The baseline entry to organized raiding
Heroic (was Normal 10 and 25)	Flexible	Its own lockout, by loot	Not stated against the other difficulty tiers	Groups of varying size, smaller ones that want to grow
Mythic (was Heroic 10 and 25)	Fixed at 20	Its own lockout, by boss, on one server, with a static ID	Not stated against the other difficulty tiers	Competitive hardcore raiders

wing is likely no faster than the second.[19] The queue's job is access to the content and its story. Separate tables, separate art and a faster reward give it that one purpose. The price is that Raid Finder gear now leads nowhere on a guild's path, and the guild raider has less reason to enter.

The calendar seam was opened on purpose. Each difficulty got its own lockout, to preserve the freedom of players who raid in several with different groups.[20] The opposite risk, the guild obliged to clear everything each week, was answered by scaling back Valor and the incentive to cap it.[21] The post cut one seam and opened another in the same breath, and the difference was whether the crossing was one the designers wanted (Figure 14.1).

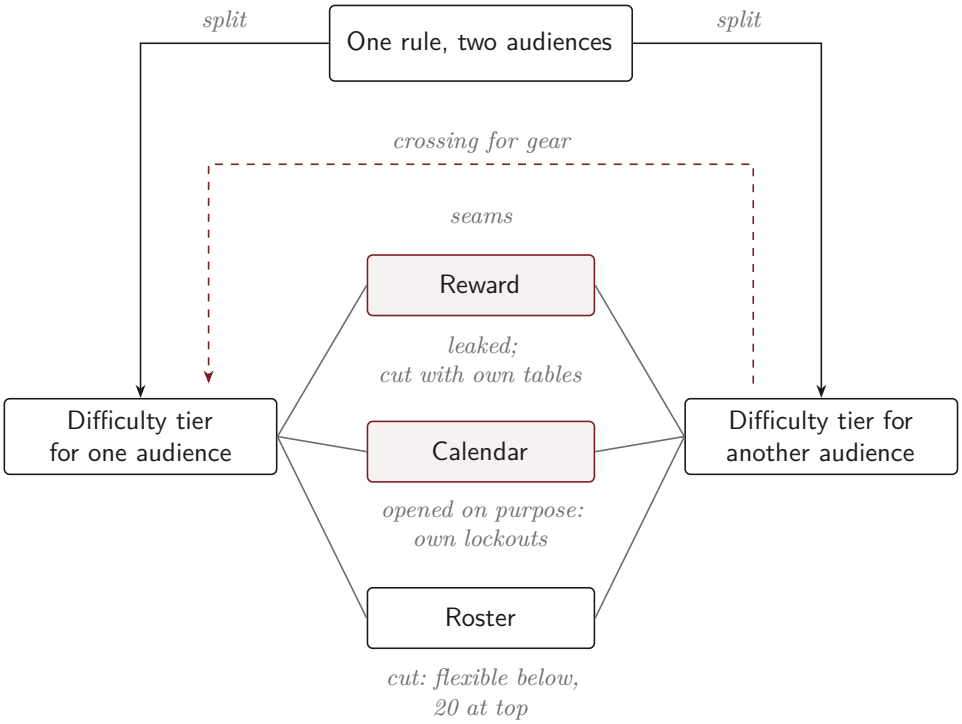


Figure 14.1. Splitting a rule into difficulty tiers moves the problem to whatever the tiers still share. The note under each seam gives what the 2014 restructuring did to it, and the dashed arrow is the crossing that leaked, guild raiders entering Raid Finder for its gear.

Cutting seams has a cost of its own. The same post wants a ladder: flexible sizes let smaller groups grow toward the fixed one, and Raid Finder is a step toward Normal.[22] A ladder needs rungs that connect, and every

rung that connects is a door of the kind just closed. Later designers put the connection in a skill: Dylan Barker describes the climb from Normal to Mythic as a rise in communication and coordination, not in damage (Chapter 12).[23] Tiers that no longer share their gear can still share a skill, and a skill is a rung that lets nobody in through the back.

14.3 The Same Split in a Character and a Card Pool

A character fights monsters and other players with one set of numbers, and the designers keep giving one mechanic two settings. Diminishing returns on crowd control is the cleanest case. In the pre-expansion notes for *Midnight*, a player fighting other players becomes immune after 2 applications from the same category, where it had been 3, while creatures no longer become immune to stuns at all; the duration instead halves with each application, down to one sixteenth of the original.[24] One rule, two tunings, sorted by who is on the receiving end.

When a number cannot be split, sharing it costs. To narrow the gap between how strong characters felt against players and against monsters, player health was roughly doubled, and creature damage and healing had to rise to pay for it.[25] When two groups want opposite things from one feature, the game director's own options are to please one group or find middle ground. Flying, once sold for gold and later ruled out, became something to unlock, a compromise he grants is perfect for neither side.[26] That is Pardo's warning about compromise, given as a choice made anyway.

Magic: The Gathering split the card pool where *World of Warcraft* split the encounter. Once Richard Garfield's early assumption that a player would own few cards gave way, and no adjustment of new cards could compete with the oldest, the answer was formats: Type 1, which allows everything under a banned and restricted list, and Standard (then called Type 2), limited to recent sets. Over time newer players drifted to Standard while Type 1 kept its loyalists.[27] Two formats work as two difficulty tiers, and the population divided by commitment and by what it already owned. Chapter 25 follows the second format into rotation.

Every format draws on one pool of printed cards, a seam. Commander, with 40 life and several opponents, undid strengths of white and red, and designers are rethinking those colors' cards.[28] A card tuned for one table is played at the other. *Magic* also has a device the raid lacks, the lenticular

card, whose second reading is invisible to the first reader, so one object serves both ends of the spectrum; Chapter 21 finds its limit.[29]

14.4 A Menu Works When Each Player Prefers Their Own Item

The difficulty tiers of a raid are a menu, and each player picks the item meant for them. Chapter 21 called the same device segmentation when rarity sorted *Magic* players by how many packs they had opened. A menu works only if each player prefers the item written for them. The guild raiders in Raid Finder are the case in which one item was worth taking by the wrong player, and the repair made it worth less to that player without making it worth less to the one it was for.

The cost of a leaking tier is the externality of Chapter 13 moved from one room to a whole population: a shared number tuned for one audience costs the other.

Where to calibrate is the last question. One school of design commentary argues for building from the top: make the game sound at the highest level of optimized play and let everyone else reach whatever share of it they reach.[30] The 2014 structure gives that argument its own tier, a fixed roster tuned for the most organized groups, with the reach downward handled by other tiers and not by the same number. *Magic*'s designers reach a neighboring conclusion: in an internal poll of rare cards, they usually prefer a card rated half very low and half very high to one rated a uniform seven out of ten, because strong impressions matter more than avoiding dissatisfaction.[31] Neither recommends the middle setting.

The three ideas fit in one line, a framework for asking where a tier will leak that returns no number.

Design equation (a conceptual framework, not a law)

$$\text{pressure of one audience on another's difficulty tier} \approx \text{value of what the difficulty tiers share} \times \text{how much the first audience wants it} \div \text{cost of crossing}$$

Raid Finder before 2014 scored high on every term: shared gear, wanted by guilds, and a crossing that cost one queue with no lockout. The change cut the first term to almost nothing for a guild. The separate lockouts lowered the last term on purpose, because there the designers wanted the crossing.

14.5 When One Rule Serves Both

A single rule can serve both audiences when what it hands one does not reduce what the other receives. The rest system of 2004 rubber-banded the rate at which hardcore and casual players leveled, rewarding casual players for taking breaks; it drew an outcry and survived, a case Chapter 20 tells.[32]

Separation has a cost the designers named long before 2014. Pardo saw dungeons as a bridge by which casual players become committed ones, and the team refused to make them playable alone because solo versions would have compromised the gameplay.[33] Tiers that share nothing have no bridge. The 2014 post keeps one of a different kind, a group-finding tool for players who want to move beyond Raid Finder, noting that pick-up groups have historically led to friendships and guilds.[34] Once the gear stopped carrying anyone across, the bridge became social.

And the audiences refuse to stay sorted. When players returned to the original game in 2019 they met the same talents and the same escalating cost of changing them, yet they farmed gold and switched builds between raiding and fighting other players in a way the players of 2005 rarely did. Hazzikostas, by then game director, put it in one line: "Same system, same talents, the level of knowledge has risen, community expectations have changed."[35] Seeking out the solved way to play is now expected at all levels,[36] and the leading teams' tools trickle down to slower groups (Chapter 15).[37] A tier's population is a moving target, and the tuning written for it ages even when no number in it changes.

The Designer's Test

1. Which audience was this number set for, and who else has to live with it?
2. If you split the rule into difficulty tiers, what do they still share: a reward, a calendar, a roster?
3. Is anything in the tier you built for one group worth taking by the other, and what would taking it cost them?
4. Where do you want players to cross between tiers, and have you made that crossing cheap?
5. How will you know who is in each tier a year from now?

Notes

1. Blizzard Entertainment. Raid Finder Q&A. `worldofwarcraft.blizzard.com`, n.d., answer to question 1.
2. Blizzard Entertainment, answer to question 2.
3. Blizzard Entertainment, answers to questions 5 and 6, and the last answer.
4. Ion Hazzikostas. Dev Watercooler: Raiding Azeroth Part 3: Warlords of Draenor. `worldofwarcraft.blizzard.com`, 2014, section "Raid Finder", paragraphs 2 to 4.
5. Hazzikostas, section "Raid Finder", paragraphs 3 and 4.
6. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "Why MMOs Are So Difficult to Create".
7. Mark Wallace. Report of Rob Pardo's keynote on World of Warcraft at the Austin Game Conference. Game Developer, 2006, paragraphs 1, 2 and 5.
8. Wallace, report of Pardo's Austin keynote, paragraph 7.
9. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., answer to "What didn't work out as planned?", heroic difficulty paragraph.
10. Blizzard Entertainment, opening line.
11. Street, "Cataclysm Post Mortem", answer to "What didn't work out as planned?", heroic difficulty paragraph.
12. Staats, *The WoW Diary*, section "November 2004: Open Beta".
13. Hazzikostas, section "Warlords of Draenor (2014)", paragraph 3.
14. Hazzikostas, section "Raid Finder", paragraph 4.
15. Hazzikostas, section "Raid Lockouts", paragraph 1.
16. Hazzikostas, section "Raid Lockouts", paragraphs 2 and 3.
17. Hazzikostas, section "Warlords of Draenor (2014)", paragraphs 1 to 3.
18. Hazzikostas, section "Warlords of Draenor (2014)", paragraphs 2 and 3.
19. Hazzikostas, section "Raid Finder", paragraph 2.
20. Hazzikostas, section "Raid Lockouts", paragraph 1.
21. Hazzikostas, section "Raid Lockouts", paragraphs 2 and 3.
22. Hazzikostas, sections "Warlords of Draenor (2014)", paragraph 3, and "Raid Finder".
23. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, last Barker quotation.
24. Blizzard Entertainment. Midnight Pre-Expansion Content Update Notes (Diminishing Returns developers' note). Blizzard News, n.d., section "Diminishing Returns", developers' note.
25. Blizzard Entertainment. Dev Watercooler: Building Healthy Gameplay. `worldofwarcraft.blizzard.com`, n.d., section "Player Health and Resilience".
26. Steven Messner. Blizzard: There are 'things to learn' from the mistakes of Battle for Azeroth. PC Gamer, 2019, answer on two types of players.
27. Mark Rosewater. Eternal Formats of the Spotless Mind. Making Magic, `magic.wizards.com`, 2016, paragraph beginning "Which brings us to 1995" and the two that follow.
28. Mark Rosewater. Why So Many Restrictions? Making Magic, `magic.wizards.com`, 2024, section "Okay, Richard built plenty of restrictions into the game".
29. Mark Rosewater. Lenticular Design. Making Magic, `magic.wizards.com`, 2014, Rules 4 to 6.
30. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 13, part A, at 11:29.

31. Mark Rosewater. Twenty Years, Twenty Lessons, Part 2. Making Magic, `magic.wizards.com`, 2016, section "Lesson #11".
32. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
33. Wallace, report of Pardo's Austin keynote, paragraphs 5 and 8.
34. Hazzikostas, section "Raid Finder".
35. Ion Hazzikostas. Dragonflight Group Interview with Ion Hazzikostas - Design Goals & Lessons Learned, Player Housing, Dracthyr. Wowhead News (post by Squishei), 2022, answer on Classic's influence.
36. Hazzikostas, answer on how players play.
37. Jaydaa, report of Barker interview, section "ADDONS & THE TRICKLE DOWN EFFECT".

Part IV

Players Against the Designer

Chapter 15

Players Will Find the Tell

The Prophet Skeram is the first boss of the Temple of Ahn'Qiraj. Chapter 9 covered his mind control and Chapter 11 why his split is tied to health. At 75, 50 and 25 percent he splits into himself and two illusory copies, and the raid has to work out which of the three is real.[1]

The designer who built him wanted one way to tell them apart: the copies would have half the health.[2] Under even damage a copy's bar falls faster than the original's, so the raid would spread its damage across all three, watch, and call out over voice chat which two were dropping quicker.[3] A wrong guess had a price. Push the real Skeram to 50 percent while the two copies from the first split are still standing, and he splits again. Now there are five of him, and the raid wipes.[4]

That is not how it went. His own first version printed a line under each copy's name marking it as Skeram's pet. In testing the team found more: interface mods that read creature IDs, and localization strings that did not match. He spent the development constantly trying to disguise the real boss among the illusions, without success.[5]

Then the fight shipped, and players found ways the testing had not. A hunter who sends a pet at Skeram before the split finds the pet stays locked on the real one. Some players mana burn him, because the copies spawn with full mana and he does not.[6] Players identified the illusions immediately, and the methods went into guides.[7]

His summary is one line: "All that work and there is still an instant tell."[8]

15.1 Five Tells, and Only One Was Chosen

A ***tell*** is any difference the game shows between two states the designer wanted to look alike. Skeram's puzzle needed exactly one, health, and nothing else a player could observe.

That requirement is much harder than it sounds. A copy of Skeram is not a picture of Skeram. It is a separate creature, created at a moment by a piece of script, in a game where dozens of other systems keep their own records about every creature in the room. Sort the tells by where they came from. The label was the encounter's own. The creature IDs came through the interface, the strings through translation, the pet's lock from another system's memory of its target, the mana from a resource the copies and the original did not start with equally.[9] One of the five is the encounter. The other four came along with the act of making a second object, from systems built for other purposes that had no reason to know Skeram's identity was a secret.

This is ***information leakage***. A fact about the game's state does not live in one place. Other things depend on it, and each dependency is a display. The intended mechanic is the only entry on that list the designer wrote.

The obvious repair is to hide the others, one at a time: take the label off, make the IDs match, fix the strings. Each repair is cheap and closes a known hole, and a team that found the label and left it in would be shipping a mistake. It still lost. The designer closes leaks in series, with a test group and a schedule. Players look in parallel: every raid that walks in runs its own search with its own mix of classes, pets, spells and mods, and whatever one raid finds, every raid can read. The designer has to remove every difference. The players need one that remains. And testing checks only the tells someone thinks to check. A pet that keeps its target is not a flaw in the Skeram script. It is a hunter's pet doing what pets do, in a room where that happens to matter.

15.2 Looking for the Hole Is Play

The designer's reaction to the pet is the sentence this chapter turns on: "If it's a real bug or an exploit, sure, we'll fix it. But players outsmarting the designer."[10] The sentence stops there. Bugs and exploits get fixed, and the hunter's pet is neither. It is also too late to fix now.

The players' side of that contest is older than the fight. In November 2003 the team opened the game to a friends-and-family alpha of around five hundred friends and partners of the staff. They found ways to cheese the game and grief one another, got past invisible walls and killer guards into areas they were not meant to reach, and filed cheerful reports about the unfinished places they found. The thrill was in the discovery and in out-

guessing the designers, and the testers felt they were part of the polishing.[11] Nobody in that passage is sneaking. They are playing a second game over the first, in which the opponent is the design team and the prize is finding the hole. The team stamped out the abuses as they came in.

One channel had been opened to players on purpose. The interface could be scripted, and within a week of the public beta in March 2004 players were writing mods: transparent windows, tic-tac-toe, and a program that found the quests a player shared with the party, one of the first fan ideas the team put into the default interface. The same week produced the first banned mod, a combat log script that whispered to itself and choked the chat channel's bandwidth.[12] A scriptable interface is a license to read the game's state, and in one week it yielded a feature and a ban. Rob Pardo drew the line later: mods that play the game for the player are bad, the rest are fine, and the harmful ones get broken in the next update.[13] The same interface code let players build a marketplace plug-in that routed around the designers' preference for trade in person, a case Chapter 16 tells.

Search asymmetry is the name for the arithmetic. The team compared rule-making to writing legislation, because rules had to curb exploits by outguessing millions of players, and one loophole could trivialize content.[14] A handful of designers check for tells. Everyone who meets the boss checks at once, and they publish.

If the search cannot be won after the fact, the cheapest moment to handle a tell is before the feature exists. The team did that twice: John Cash ruled out offline adventuring because the software on a player's computer can always be hacked (Chapter 24), and Eric Dodds answered a bounty pitch with the ways players would cheese it (Chapter 3).[15] Both assumed the search would happen and designed for the searchers.

15.3 What to Call a Find

Every difference the system can show is a signal, whether or not the designer meant to send it.

Once a tell is found, the harder question is what to call it. Three questions sort the four common labels. Does the player still take the test the designer set? A route nobody planned that still does what the content demands is clever play; one that skips the demand and collects the reward is an ***exploit***, the only sense in which this chapter uses the word. Was the behavior intended at any level? A bug is something the code does that nobody on the team would defend as a decision. Does anyone enjoy it? A ***degen-***

Table 15.1. Five discoveries from this chapter, sorted by label, by who else paid, and by what the designers did.

What was found	Label	Who else paid	What the designers did
A hunter's pet stays on the real Skeram	Clever play: the raid still has to fight the split	Nobody outside the raid	Left it alone
A taxi dismount into a closed zone, then a vendor resale	Exploit: the gate on those weapons was skipped	The buyers, and the economy	Removed the dismount
A spot where a monster cannot reach its attacker	Bug, and a degenerate strategy if it stays	The economy, and the user's own leveling	Named pathing code the remedy
A combat log script that whispers to itself	Bug, in a player's code	Everyone on the chat channel	Banned the mod
A marketplace plug-in for anonymous trades	Clever play against a preference	The designers' preference	Built the auction house

erate strategy is a legal, efficient way to play that makes the game worse for the person using it. A fourth question cuts across all three and decided more outcomes than the labels did: who pays for the discovery besides the players who use it.

The first row needs care. The narrow test, reading health bars and talking about them, is skipped by a raid with a hunter. The wider one is not: the raid still has to kill the right creature in the right order across three splits. The taxi resale and the unreachable monster are alpha cases Chapter 3 tells. The resale earns the word exploit by its effect, since the zone was closed to gate those weapons behind levels of play and the route sold them to other people.[16]

Read the last two columns together. The pet costs nobody outside the raid, and it was left alone. The resale and the unreachable attacker put goods into an economy every other player lives in. The whispering script

took bandwidth from everyone on the channel. Those three were removed, named for a fix in the pathing code, and banned. A find gets celebrated when the finder bears its consequences and gets closed when strangers do.

Where the blame goes follows from that. The designers were amused at the artist's ingenuity, and the diary compares developers who blame players for ruining a game to dog owners blaming their pets for eating scraps.[17] Greg Street, writing about hunters who stacked cooldowns in arenas, put class balance the same way: the team did not blame players for using the tools it gave them, and blamed itself.[18]

15.4 Cards Have No Creature ID

Magic: The Gathering prints every card's text on its face, so a card has no hidden fact to leak, though the hand and the deck are hidden and every play shows a little of them (Chapter 7). What players find on a card is a use nobody priced. Lion's Eye Diamond was a narrow card that became a key piece of a top deck at Pro Tour Rome in 1998, because a card's worth shifts as the field of decks around it shifts.[19] The error runs the other way too. R&D expected a cycle of cards called Shrines to matter in multiplayer games, and they did not.[20] The designers' estimate of a card is one search, run before printing by a small group. The players' is another, run afterward by everyone, and the two disagree in both directions.

Some finds are structural. *Magic* divides its effects among colors, and that division limits what one deck can mix. Artifacts walk around the limit: they tend to be colorless, so any deck can run them, and they combine into packages that are hard to stop. Artifact blocks are often the ones cited for bans.[21] Chapter 4 reads that as a point about constraints. Here it is the card-game form of the hunter's pet, a general-purpose part passing through a wall built for a specific purpose.

Magic's designers also close doors before players reach them. Clauses such as once per turn and sorcery-speed timing are a core tool, and controlling when and how often an effect happens makes better designs, as with Rootwalla in Chapter 4.[22] Such a clause is leak control at design time: it bounds what any later card can do with this one, including cards nobody has thought of yet. The cost is that players dislike reading the restrictions, which R&D now states bluntly, like vegetables once hidden in the dish and now served in plain view.[23]

15.5 When the Search Runs on Scripts

The first week of mods has a present-day sequel. Teams racing to be first in the world through a new raid hired programmers to write WeakAuras, custom scripts that countered a boss's mechanics, and the encounter designers then built mechanics with those programmers in mind. The contest trickled down to slower raid teams and into Mythic+ dungeons, to the point summed up in one line: "You probably shouldn't need coders to raid." The response was a restriction on the combat data add-ons may use.[24]

This is Skeram's contest with one change: the players' side of the search is automated, and a script reads a tell faster than anyone can call it on voice. Add-ons could solve a fight's strategy but could not steer a character out of the fire, so the difficulty that remained fell on dexterity and reflex.[25] When one layer of a challenge can be computed, the challenge moves to the layer that cannot.

> **Second-Order Effect: Private auras**
>
> Before restricting the data, the encounter team tried hiding it. A private aura kept an effect on a player out of add-ons' reach, with the aim of making raids more interactive and less computational. One of the leads later called the feature an experiment from the start and said private auras can be worse than the problem they are trying to solve.[26] A state hidden from the tool was also hidden from the players who had come to rely on tools to see it, so the fix aimed at the scripts landed on the raid.

Closing every leak is still the wrong lesson. The quest-finding mod supplied what the default interface lacked. Private auras closed a leak at a cost their builders judged sometimes worse. And a find can be evidence: when a mechanic gives players too little time or feedback to respond, they turn to add-ons, and the same leads say some recent fights needed more of both.[27] A tool is often how players report that the game's own display has failed them, and breaking the tool silences the report. Skeram's later history makes the point from the far side. Asked about fights built on communication and uneven information, a current lead named Prophet Skeram as the precedent and expected more such mechanics as computational add-ons recede.[28] The fight whose puzzle leaked is the model for the puzzle.

The rule that a find is closed when strangers pay has one hole, the marketplace plug-in, which cost the designers a preference they cared about

and was adopted anyway. It does not say what a designer's preference is worth once players have voted against it.

The definition has one limit. A tell is a difference nobody meant to show, which assumes nobody on the other side is arranging what the searcher sees. Against a person, somebody is. The Distraction Makers channel calls it donkey space: a player plays badly on purpose, so the opponent cannot separate weakness from pretense and must then wonder whether the weak play was the bait. Neither randomness nor a system produces this.[29] Skeram's copies had no say in what a pet or a mana bar showed. Against a mind the tell can be manufactured, and the search never closes.

The Designer's Test

1. What do you want hidden, and what else in the system changes when that fact changes?
2. Which of those changes can a player, a pet, a combat log or an interface script observe?
3. If players find a route you did not intend, do they still take the test you set?
4. Who pays for the discovery besides the players who use it?
5. What would hiding this cost the players who were never going to look?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V3 at 1:28.
2. Candlesan, commentary, V3 at 1:34.
3. Candlesan, commentary, V3 at 1:45.
4. Candlesan, commentary, V3 at 1:51.
5. Candlesan, commentary, V3 at 1:59 and 2:22.
6. Candlesan, commentary, V3 at 2:33.
7. Candlesan, commentary, V3 at 2:27.
8. Candlesan, commentary, V3 at 2:47.
9. Candlesan, commentary, V3 at 2:05.
10. Candlesan, commentary, V3 at 2:51.
11. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "November 2003: Friends-and-Family Alpha".
12. Staats, *The WoW Diary*, section "March 2004: Public Beta 1.0".

13. Vincent Diamante. Report of Rob Pardo's talk on the Blizzard approach to multiplayer design, Game Developers Conference. Game Developer, 2008, section "Dealing With The Player".
14. Staats, *The WoW Diary*, section "Gameplay".
15. Staats, *The WoW Diary*, sections "March 2001: My First Six Months" and "Announcement at the ECTS".
16. Staats, *The WoW Diary*, section "September 2002: Internal Alpha 1.0".
17. Staats, *The WoW Diary*, sections "September 2002: Internal Alpha 1.0" and "October 2002: Still Unanswered Questions".
18. Greg Street. Dev Watercooler: Mists of Pandaria PvP. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "Beastmastery Hunters".
19. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, section 4.
20. Rosewater, "When Cards Go Bad", section 7.
21. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section "Kaladesh, Lessons" and "Wrapping Things Up".
22. Mark Rosewater. Why So Many Restrictions? Making Magic, `magic.wizards.com`, 2024, section "But Magic has been fun for many years".
23. Rosewater, "Why So Many Restrictions?", section "Does it bother you that some of the new restrictions feel bad to read?".
24. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section "ADDONS & THE TRICKLE DOWN EFFECT".
25. Jaydaa, report of Barker interview, section "ADDONS & THE TRICKLE DOWN EFFECT".
26. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer beginning "That's awesome. I know you guys have been responding".
27. De Sousa and Sanders, interview, answer beginning "It seems that some mechanics don't give players enough time".
28. De Sousa and Sanders, interview, answer beginning "We haven't really explored the communication or information asymmetry mechanics".
29. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 6 at 27:34; episode 22, part B, at 7:46.

Chapter 16

Players Are Better at Finding Problems Than Solutions

When Blizzard decided to replace the talent trees of *World of Warcraft*, the designer responsible built his announcement out of the complaints. Greg Street, the game's lead systems designer, published a long post about the overhaul planned for the next expansion. Each heading was something players had been saying. Under each heading was his answer.[1]

The first heading was that players now had fewer choices. He did not dispute the count. In the old trees, he wrote, a player picked the talent wanted from among talents that held no interest, which is a multiple-choice test, and an easy one. A Retribution paladin simply takes the damage option. Make that player choose among healing talents with no way to trade healing for damage, and the choice becomes real. The new design would have fewer choices and more that mattered.[2]

Under other headings he agreed. A talent point at every level had worked when a character had sixty levels to climb and did not work at ninety. Abilities would now arrive automatically, and he did not claim that solved the problem.[3] Some new talents only mattered when fighting other players, since snares, crowd control and defensive cooldowns do little against a raid boss, and he called the monotony of outdoor encounters a problem with the game, not with the people pointing at it.[4]

At the end came the position that governs the rest. Feedback informs the decision and does not make it, and a good game cannot be designed by popular vote.[5]

Count what happened. Several complaints were granted in full, in public, by the person with the authority to act on them. One was contradicted. None of the remedies was adopted, and the decision to replace the trees stood.

16.1 Every Complaint Carries Two Messages

Feedback arrives as one message and contains two. The first is a report: this feels like less than I had, leveling feels empty. The second is an instruction: give the trees back. They come fused. A player does not write that a character's growth has stopped feeling like growth and then, separately, submit a proposal. A player writes that the new talents are a mistake. Volume and eloquence attach to the instruction far more readily than to the report, because an instruction is an argument and a report is only a feeling.

Two policies treat the message as one thing. Obedience takes it as a work order: restore the trees. Deafness discards it, because forums always complain and the people who post are a sliver of the people who play. The game's own development diary records that attitude in the building: when an earlier system drew fire, the fans on the forums went berserk, as per usual.[6] If outbursts are routine, any single one tells the designer nothing.

Obedience runs into a constraint the players cannot see from their seat. A tree of forty or fifty talents that are all fun and all balanced cannot be built, because of the number of combinations they form with one another. In the trees being replaced, the legitimate choices in many specs came to between one and four points.[7] A player asking for the big tree back is asking for something every player could enjoy and no designer could deliver. The combinatorial bill is paid at a desk the player has never seen.

Deafness throws away a measurement. The leveling complaint was true, and so was the complaint about talents that only matter against other players.[8] No internal tool reports that the stretch between two levels has gone flat. A person has to feel it and say so. The complaints in the post split cleanly: where players described an experience, the designer agreed, and where the complaint implied a design, he did not.

16.2 What Players Asked For and What They Did

The same post shows what players did with the choices they were defending. Builds that departed from the standard one were usually weaker. A build copied from a website, Street replied to one objection, is not skill.[9] Players defended a freedom that most of them exercised by following a guide. That is no charge of hypocrisy. Copying a solved build is the sensible response to a system with a best answer in it, as Chapter 1 argued about every efficient response in this book. What players asked for and what they did pointed in different directions, and only one was visible in the complaint.

This is the gap between ***stated preference*** and ***revealed preference***, between what people say they want and what their choices show. Chapter 1 read the revealed incentive off what a rule pays for. Revealed preference turns the same reading on the players. Neither kind of evidence outranks the other by default. Behavior is harder to argue with and silent about feelings, and a statement can report a feeling that leaves no trace in behavior, as the flat stretch between two levels did.

The rest system Chapter 20 tells is the same split in another form: the stated objection was the limit on experience, and the limit stayed once it was presented as a bonus.[10] So is the first week of raiding. When the open beta brought the raid dungeons to the public, the first trash monsters wiped groups again and again, and many fans concluded the raids were buggy, or that the level cap would rise, or that groups would later be allowed more players.[11][12] The symptom was accurate: groups were dying at the door. The cause was poor armor and no buffs or flasks, and each of the three guesses implied a different change to the game.

Some feedback arrives with the prescription already removed, because nobody speaks it. In the first test of Alterac Valley (Chapter 7), developers who had died ran the wrong way because the mini-map did not yet show their allies. The diarist's verdict on them: "they were the perfect playtesters."[13] None had to write that the map needed ally markers. Where they ran said it.

16.3 Believe the Report, Audit the Remedy

Mark Rosewater reached the same conclusion from twenty years of designing *Magic: The Gathering*. His nineteenth lesson is that players are good at recognizing problems and poor at solving them, so their feedback should be used to find the problem and the fix they propose handled with caution.[14]

Believe the report. Audit the remedy.

The reason is the seat. The player is the only instrument that can measure how a rule feels. The same player has the worst view in the house of what the rule holds up: the other classes, the other formats, the production schedule, the combinations. A report needs one seat. A remedy needs all of them. This is ***diagnosis versus prescription***: a complaint makes two claims of separate reliability, the report drawing on what the player has lived and the instruction on a model of the whole game the player has had no chance to build. The Distraction Makers channel separates a fan's knowledge, an internalized sense of a game absorbed from long play, from

a critic's, a technical grasp of how the thing is made, and holds that a designer needs both.[15] Long play supplies the first in quantity and does not teach the second, so the report stays the more reliable half even for a player shown every constraint.

That gives a way to take a complaint apart before acting on it. Split it into five parts: the ***symptom*** the player reports, the incentive underneath it, the constraint the designer works under, the players' solution, and the designer's. Table 16.1 applies it to three complaints.

Table 16.1. Three complaints taken apart into five parts. Cells are drawn from the sources cited in the text.

	Fewer talent choices	Reset the rating	Raids wipe at the door
Symptom	The new design feels like less	Top titles look as if they can be had without playing	The first trash kills the group, repeatedly
Incentive underneath	A tree with a best answer gets solved and copied	A high rating, once reached, is safest left alone	n/d. The diary names a cause: poor armor, no buffs or flasks
Constraint	Forty or fifty balanced talents cannot be built	The rating exists to make matches, not to keep score	The designer's math (details n/d)
Players' solution	Keep the trees (implied by the complaint)	Reset the rating each season	None stated. Three guesses: bugs, a higher level cap, larger groups
Designer's solution	Fewer talents, each meant to matter	Let ratings grow while a team keeps playing	None. No retuning is recorded

Read the columns downward. In each, the first row is something the player measured correctly, and the fourth was written without sight of the third. The fifth differs from the fourth in every column: a different design, a different lever, no change at all.

The middle column is the cleanest, because there the designer did the whole decomposition himself. Teams in fights between players carried a matchmaking rating, the number the game used to pair them with opponents of similar strength, and many players had asked for it to be reset each

season. The fear underneath was that teams with a high rating could sit on it, stop fighting, and still collect the season's top titles. If a title is awarded on a number and playing can only put the number at risk, the best move at the top is to stop. Street granted the fear some legitimacy and named the constraint: the rating is a tool for making matches, never meant as a score, and resetting it is the wrong answer. His remedy was to change the team rating formula so ratings grow while a team keeps playing, which means an idle team at the top is overtaken by teams that play and win.[16] The players' diagnosis survived intact. Their prescription was declined, and the alternative aimed at the incentive they had identified (Figure 16.1).

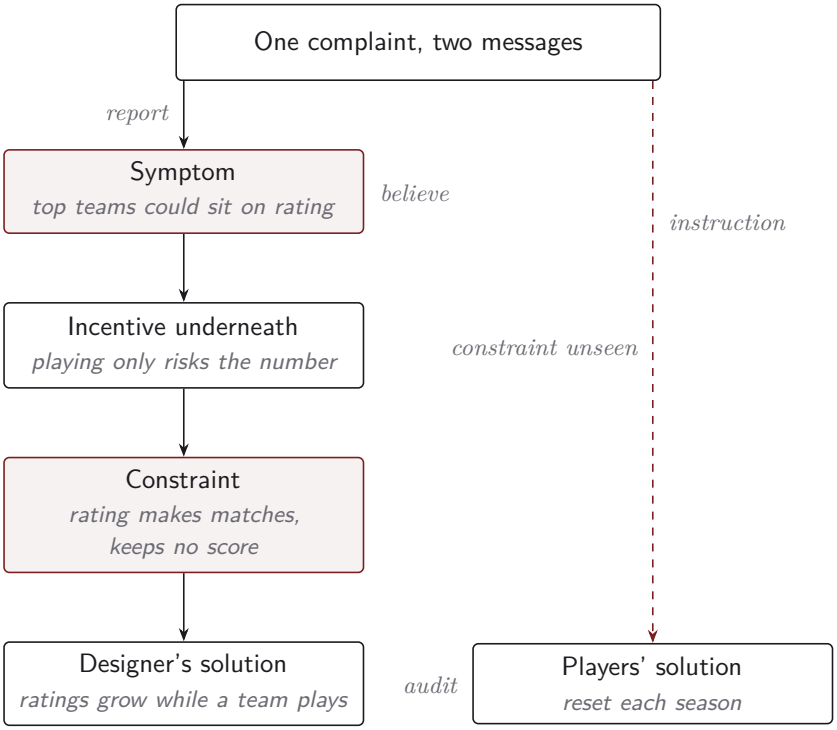


Figure 16.1. One complaint taken apart, with Street's post on the rating reset as the instance. The players' remedy is written without sight of the constraint.

Magic's version begins with a reader's letter about weak rare cards.[17] The symptom is real: opening a rare and finding nothing to do with it feels bad. The implied fix, make every card good, is the one thing the system cannot bear. A card is strong or weak only beside the cards around it, and Volcanic Hammer looks poor because Lightning Bolt does the same

job better. Of roughly 1,500 cards in the pool, 300 to 400 would see play, between a fifth and just over a quarter, so most cards must earn their place somewhere else, and raising every set's power to keep its cards wanted would in time break the game.[18] Players also rate each card by their own needs, Limited, multiplayer, flavor or casual, so bad very often means not for me.[19]

The proposal is correct for one pack opened by one person. It is local optimization in the sense of Chapter 1, a real problem solved well at a horizon smaller than the designer's, and here the horizon is one seat. Each proposal optimizes one seat, and the constraint lives in the sum over seats. A design aimed at every seat at once gets the result of Molten Sentry, the coin-flip card of Chapter 6, which pleased neither the players who like variance nor those who dislike it.[20] A vote sums fixes with nobody holding the constraints, and Blizzard had tried something close to it on itself: Nomad, the project Chapter 22 tells, ran on democratic design and ended pleasing no one, and the team that built *World of Warcraft* worked instead under a representational democracy in which designers heard every argument and owned the decision.[21]

16.4 How Hard to Act on a True Report

Trusting the symptom leaves open how hard to act on it, and the talents post admits a miss.

> **Second-Order Effect: The cautious fix**
>
> The trees Street was replacing had been redesigned one expansion earlier. That redesign was restrained on purpose, to avoid shocking players, and it did not fix the underlying problems. The drastic overhaul was the result, and he defended it by saying the team will not leave something at a B-minus if it believes an A-plus is within reach.[22] A careful answer to a true complaint bought one more expansion of the same complaint, then a larger upheaval than the caution had been meant to spare. Chapter 2 tells that overhaul from the players' side.

How hard to act depends as well on when. Faced with a worry that the new monk class was too weak in arenas, Street chose to change nothing and watch for a few weeks, so as not to repeat the death knight, the previous new class, which had dominated play against players and against the game alike.[23][24] Chapter 20 tells that story. Early reactions to something new

mix real imbalance with unfamiliarity, and the loudest channel is not the population.

Timing sets the last limit. R&D found its complexity problem while the *Time Spiral* block was too far along to change, and the repair moved to the next block (Chapter 21).[25] A correct reading that arrives after the product is committed fixes the next product.

16.5 When the Players' Fix Is the Right One

Sometimes the players' solution wins, and in this record those are the cases where players built it. The designers wanted goods to change hands person to person, a social preference. Players used the game's own interface code to build a marketplace plug-in that moved items anonymously, and its popularity tipped the scales toward an auction house. The diarist was not surprised, and found it disconcerting that the team was changing the game because it could not stop players from implementing their own system.[26] The quest-sharing mod of Chapter 15 went the same way into the default interface.[27]

In that same auction house a typed request was refused. Players asked for proxy bidding, which would bid for them while they were elsewhere, and the team declined because the auction house was meant to be gameplay in its own right, so that life as a trader in the city stayed something a player does.[28] A solution built and adopted has already met the players and survived. It is behavior and carries behavior's weight. A proposal does not. Rosewater's first lesson is the matching case, the suspend creatures of Chapter 1, where players kept making an illegal attack until the designers wrote it into the rules.[29] His first lesson and his nineteenth agree once what players do and what players propose sit in separate columns.

The principle also flatters the designer, and stated taste is unreliable on both sides of the table. Inside the company almost nobody liked split cards, which print two small cards on one face, and both R&D and the brand team tried to remove them. Market research ranked them the most popular mechanic.[30] The death knight overshot and the cautious talent redesign undershot.[31][32]

Then there are the raids. While groups wiped at the door, developers asked Jeff whether the content was meant to be that hard. Raids looked as if they would appeal to 1 percent of the player base, one player in a hundred, and many on the team doubted they were worth building. Jeff relied on his math: "Trust me, they'll get it."[33] The refusal reads as wisdom

in hindsight. Had they emptied, the same sentence would be quoted as a designer ignoring his players. The team's own whiteboard, quoted at the head of the diary's launch day, knew it: "Vision is what some people claim to have when they've guessed correctly."[34] The principle sorts feedback into what to believe and what to check. It gives the designer no exemption from being checked in turn.

The Designer's Test

1. What exactly do players report feeling, once their proposed fix is removed from the sentence?
2. What does the rule you are defending hold up, and can the people complaining see it?
3. If you did exactly what was asked, what would break, and for whom?
4. Who is not speaking, and what are they doing in the game?
5. Have the players already built their own fix, and is anyone using it?

Notes

1. Greg Street. Seeing the Forest for the Talent Trees. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d.
2. Street, "Talent Trees", Section 1.
3. Street, "Talent Trees", Section 4.
4. Street, "Talent Trees", Section 6.
5. Street, "Talent Trees", Section 10.
6. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "May 2004: The Care Bear Game".
7. Street, "Talent Trees", Section 2.
8. Street, "Talent Trees", Sections 4 and 6.
9. Street, "Talent Trees", Sections 2 and 3, and Section 5.
10. Frank Cifaldi. Report of Rob Pardo's talk Making a Standard: Blizzard Design Philosophies, Game Developers Conference. Game Developer, 2010, section "Make It A Bonus".
11. Staats, *The WoW Diary*, section "November 2004: Open Beta".
12. Trash is the players' word for the ordinary monsters that stand between a raid and its bosses.
13. Staats, *The WoW Diary*, section "October 2004: World's End".
14. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section "Lesson 19".
15. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 14 at 10:41 and 11:32.

16. Greg Street. Dev Watercooler: Mists of Pandaria PvP. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., MMR section.
17. Cards are printed at several rarities, and a rare is one of the scarce ones.
18. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, opening letter and Section 1.
19. Rosewater, "When Cards Go Bad", Section 2.
20. Rosewater, "Twenty Lessons, Part 3", section "Lesson 15".
21. John Staats. Making WoW: Learning from the Mistakes of Nomad. Wowhead, 2019, paragraph 1 and last paragraph.
22. Street, "Talent Trees", Sections 8 and 9.
23. Street, "Mists of Pandaria PvP", Monks section.
24. Blizzard Entertainment. Blizzard Insider 42: Meet the Pandaren Monk (interview with Greg Street). `worldofwarcraft.blizzard.com`, n.d., answer on death knight lessons.
25. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "A Complex Problem".
26. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
27. Staats, *The WoW Diary*, section "March 2004: Public Beta 1.0".
28. Vincent Diamante. Report of Rob Pardo's talk on the Blizzard approach to multiplayer design, Game Developers Conference. Game Developer, 2008, non-combat PvP paragraph.
29. Mark Rosewater. Twenty Years, Twenty Lessons, Part 1. Making Magic, `magic.wizards.com`, 2016, section "Lesson 1".
30. Rosewater, "Twenty Lessons, Part 3", section "Lesson 16".
31. Blizzard, "Meet the Pandaren Monk", answer on death knight lessons.
32. Street, "Talent Trees", Section 9.
33. Staats, *The WoW Diary*, section "November 2004: Open Beta".
34. Staats, *The WoW Diary*, section "Launch Day".

Chapter 17

An Unwritten Rule Is Still a Design Decision

In July 2020 Ion Hazzikostas, the game director of *World of Warcraft*, was asked about loot in the coming *Shadowlands* expansion. Part of his answer concerned modes the game no longer offered. The reasons for moving from group loot and master loot to personal loot three or four years earlier had not meaningfully changed, he said. The old modes allowed abuses in which players lost control over their own items, because in effect "the items now belong to the guild master, or a PUG leader that has arbitrary rules," and the trouble grew worse when items were scarce.[12]

Under master loot, the rule about who got the sword was whatever the person holding the bag said it was. Personal loot ended that: the game decided, for each player, what that player received.

Two years later the game moved the other way. Blizzard's rules for Vault of the Incarnates, the first raid of *Dragonflight*, made group loot the only method of distribution, with no option to return to personal loot.[3] Each player answers each item with need, greed, disenchant or pass. Need beats greed and disenchant, and ties go to a roll. Every item has the same chance of dropping whatever the group's makeup, so a player whose class stands alone in the raid has the same chance as anyone else of seeing an upgrade.[4]

Within a few years one question, who takes the item, had been answered by a leader, by the code, and by a roll each player decides how to enter. Each answer was a designer's decision.

17.1 Three Places to Settle a Behavior

A multiplayer game runs on behaviors no rule states. Who takes the drop when several people killed the monster. Whether to kill a player far weaker than you. How long to think while three other people wait. Whether to

bring your strongest deck to a table of friends. Each has to be settled for the game to work, and the rulebook can be silent on all four.

The designer has three places to put the answer. The first is the code: the act becomes impossible, and nobody has to decide anything. The second is a written rule with someone to enforce it, a referee or support staff, so the act stays possible and becomes forbidden. The third is nowhere, and then players supply an answer of their own. That answer is a ***norm***: an expectation no rule states and players enforce on each other. The place where a behavior is settled is its ***holder***, and the three holders match three verbs: cannot, may not, should not.

The third holder is the cheapest to ship, because nothing has to be built, and the only one that changes without a patch, which helps when the designers guessed wrong and hurts when the players drift. Table 17.1 places seven behaviors from this chapter.

World of Warcraft's early team reached for the code wherever it could. *Magic*, played with cards the players bring, began with deck size and copies left open and moved both into written rules. The loot row is the only one that changes holder more than once.

17.2 The Code Where Possible, Custom Everywhere Else

The team had a stated policy. John Staats's diary records it beside a rejected alternative: another studio's game engineered its community with incentives for good conduct, which he found a bizarre priority, and his own pitch for a bounty system had died on Eric Dodds's objection that players would cheese it (Chapter 3). An incentive aimed at conduct was one more thing to exploit. Chapter 3 quotes the first half of the team's answer, that it simply prevented the ability to grief. The second half is this chapter's: "If our designers could prevent griefing and minimize reasons to argue, we were laissez-faire about social behavior."[5]

What could be prevented went into the code. A trade both sides must confirm cannot be swapped at the last second, so no custom against swapping has to exist, and the diary counts confirmation buttons and no-drop rares among the details that kept the game fun by preventing griefing, stabilizing the economy and keeping promised rewards.[6] When the artist of Chapter 3 dismounted from a taxi into a closed zone and killed colleagues with campfires, the dismount was removed and Dodds concluded that player fires should do no damage.[7] No warning, no punishment: the act stopped

being possible. The second holder barely appears. One open question of October 2002, listed as one of the more amusing, was to sentence abusive players to an exile server the GM staff would not support instead of cancelling their accounts.[8]

Whatever could not be prevented went to the third holder by default, and there the policy fails at its condition. A norm that the rules pay players to break is a tax on the polite. The first company-wide alpha, which Chapter 1 tells, let anyone loot any corpse; the melee players got everything, and everyone's goal became to level out of the crowd and play alone.[9] These were colleagues in one building, 61 of them at the Friday peak, which the email called an all-time high, a group as able as any to keep a custom of leaving other people's kills alone.[10] No custom formed. A tester who declined to loot went without, and the one who took the item was paid by the rule that made courtesy expensive. Open looting had the shape of Garrett Hardin's tragedy of the commons, a shared stock each user is tempted to draw down, with restraint a private cost and its benefit shared.[11]

Whatever the rules leave open, a norm fills or nothing does, and the designer who leaves it open has chosen whoever ends up settling it.

Fighting between players could not go into the code without removing something the game was for. The diary files it beside instanced dungeons as one of the team's religious issues, since it went hand in hand with griefing and the question was where to draw the line, and many developers wanted no servers for it at all, fearing a poisonous environment.[12] The answer was a choice of server. By July 2004 the diary reports that three of four Koreans preferred the servers with open fighting while the American player base split evenly, three in four against one in two.[13] Where no single line suits everyone, each player chooses the rule set to live under, and the norm question becomes a menu.

The open server shows what the third holder produces. When the second public beta in June 2004 opened one, the Alliance went live outnumbering the Horde two to one, and undead players could barely reach new zones without being ganked before they could bind there.[14] Then the Horde rallied, and the game "was finally playing like a war!"[15] One open rule produced the ganking and the war.

A norm also belongs to a place, and a game ships to many. The team's briefing on the Asian market in December 2002 said that games with nonconsensual fighting between players received adult-only ratings, that in Korea real police responded to griefing, and that because companies had to

keep storefronts there by law, players would fly across the country to complain in person.[16] A studio that leaves conduct to custom ships a different game to each country without deciding to, because the norm that fills the gap is the local one.

17.3 Moving Loot Into the Code

Personal loot puts the allocation in the code and leaves the player one act: an owner can trade an item away. With random upgrades no longer interfering with tradability, Hazzikostas said, personal loot hands out upgrades more consistently.[17] Consent replaces discretion. Under a leader, a player received what the leader chose to give. Under personal loot, a player gives away only what that player chooses to give.

The game had already written etiquette into the roll for groups of strangers. The Raid Finder, whose structure Chapter 14 sets out, gives priority to a need roll on an item that fits the player's primary role, lets a player win loot from a given boss once a week, and gives the raid leader no extra power to invite.[18] In a standing group those would be customs. In the queue they are properties of the code, and code that replaces a custom has to guess the custom's exceptions.

> **Second-Order Effect: The exception the code did not know**
>
> Personal loot was extended to the game's old dungeons and raids because it let the team stretch leveling dungeons across more levels. Players who clear old content for the look of its items found it worse, and a staff reply granted that it was not a great experience for those who clear older content for transmog.[19] The answer was a further rule, a legacy mode for content more than 10 levels below the character, in which bosses drop as many items as they would for a full party.[20] A leader with discretion would have handed a lone collector the whole table unasked. The code needed a clause keyed to a level gap. Removing master loot from current raids also complicated how guilds handed out certain untradeable mounts, which was hotfixed.[21]

The roll of 2022 holds half the decision. The code decides who wins among those who pressed the same button. Each player decides which button to press, and the text never says when need is the right answer. It accepts one consequence in advance: a bow may drop for a group with no

hunter, presented as a reward for bringing a diverse set of classes.[22] Under the roll, part of the allocation is held again by what the group thinks is proper. A decision taken into the code can come back out, so the principle says what a designer chooses by leaving a gap, not that the code is always the better holder.

17.4 Magic Writes Its Norms Down, Except One

Magic's first rulebook, in the summer of 1993, left deck building almost open: a minimum of 40 cards, no maximum, no limit on copies.[23] An expectation about people stood in for the limit. Richard Garfield assumed players would spend on the game about what they spend on an average board game, so few would own many cards and fewer still more than one or two of the most disruptive ones. Players bought boxes of booster packs.[24] Tournaments showed the trouble. The default quickly became 60 cards and four copies, and in early 1994 the Duelist Convocation wrote both into its tournament rules, from which they entered the official rules.[25] Chapter 25 reads that history as a defense against power. Here a behavior moved from an expectation to a written rule. Beside mana cost and the four-copy cap, a trading card game also has a secret third limiter, price, which keeps players who cannot or will not pay from owning the strongest cards.[26] It does a rule's work without being one.

Pace sits in the middle holder. The tournament rules put no clock on a single decision. They require players to take turns in a timely fashion whatever the complexity, forbid stalling, and let a player ask a judge to watch for slow play.[27] A counted limit per decision would be unworkable, and silence would let a slow player steer the round's clock (Chapter 8).

Commander runs on the third. It is a format players made, which spread until the company built a product for it and then one every year, and the most played casual format.[28] No rule stops a player from bringing a deck built for the most competitive tables. A norm says not to.[29] The designer controls the rules and not the culture, which passes by word of mouth.[30] The work of restraint falls on players who cannot measure what they restrain: outside the most competitive decks, a deck sits somewhere vague in the remainder, and learning whether it is truly strong would take hundreds or thousands of games.[31] The more anyone tries to define casual play, the more exclusionary it becomes, and even limits that only say what casual play is not end up prescribing a little.[32] When someone wins in a way the table dislikes, players blame the player, sometimes to screen their own failure,

though the system allowed it.[33] Ninja, in Chapter 1, names the looter in the same way, not the rule that paid him.

Persuasion marks the edge of writing. Politics at a table of four (pleading, threats, deals, pointing at someone else) sits almost entirely outside the game's systems, while a duel leaves almost no leverage, since a player must attack the only opponent there is.[34] A judge can watch pace. Persuasion among four players can only be held by norm.

17.5 How Far a Norm Holds

What separates the holders is who pays for enforcement (Figure 17.1). Code is paid for by the designer, once, plus the exceptions it turns out to need. A written rule is paid for by a referee each time someone calls one. A norm is paid for by whoever objects, every time, and only if that person can make the objection cost something. Under master loot, the people most likely to object were the people without the bag.

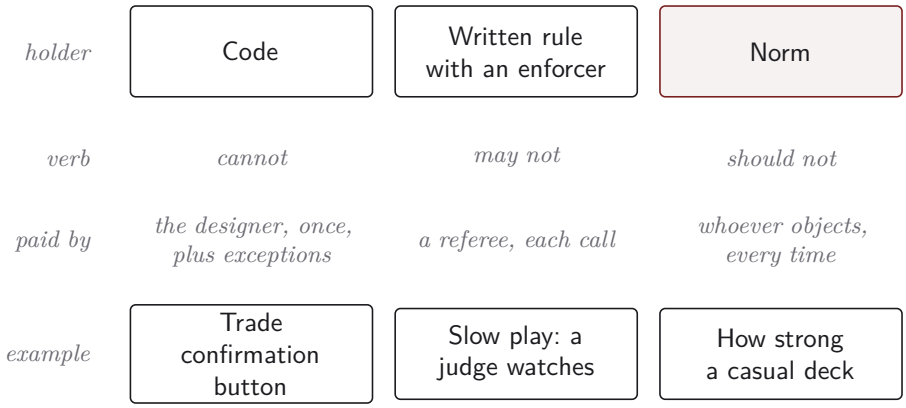


Figure 17.1. Three places a designer can put a behavior, and who pays to enforce each.

A settled practice, a table that plays below its strongest decks or a group that agrees what need means, solves a coordination problem of the kind Chapter 12 took from Thomas Schelling.[35] Nothing but expectation holds it in place, so it lasts while each player expects the others to keep it. Chapter 1 showed that the pressure a rule produces depends on the share of players who optimize. A norm is strong in the same proportion: it survives a few optimizers and fails once enough of them arrive.

Design equation (a conceptual framework, not a law)

$$\text{norm strength} \approx \text{visibility of a breach} \times \text{cost others can impose} \times \text{chance the same people meet again}$$

The leader's bag fails the first two terms. Under arbitrary rules there is no standard for a breach to show against, and the one person able to impose a cost is the person holding the bag. A queue of strangers fails the third, and the Raid Finder writes its manners into the roll. A guild will keep a custom a queue will break.

A designer can also want a norm and lose it. The team preferred trade in person and got an auction house (Chapter 16). Rob Pardo accepted the friction of slow travel by flight path, and the frustration players voiced about it, for the sake of the social connection he said such games are about.[36] In the diary's first entry the team argued over private dungeons, whose biggest problem was that they were antisocial in what was meant to be a social game, in meetings that ran until people broke into small groups arguing about the genre, sometimes until midnight.[37] Each convenience removed an occasion on which players had to deal with each other, and with it a place where a norm could form.

The Designer's Test

1. Which behaviors does this system depend on that no rule states?
2. For each, who enforces it today: the code, a referee, or whoever in the group holds the most power?
3. Does any rule you wrote pay players to break the custom you are counting on?
4. If you move this behavior into the code, what exception did the custom handle that the code will not?
5. Will the people in this group meet again, and does your answer to the second question assume they will?

Notes

1. Ion Hazzikostas. Ion Hazzikostas on Shadowlands End-Game Loot - Utility Legendaries and 'Loot is Loot' Droprates. Wowhead News (transcript by Perculia), 2020, second answer.
2. A PUG is a pick-up group, assembled from players who do not usually play together. Under master loot the items a raid earned were, in effect, the leader's to give.
3. Unsigned. Group Loot Only in Vault of the Incarnates - No Personal Loot in Dragonflight Raid. Wowhead News (post by Anshlun), Blizzard text, 2022, Blizzard block under "Group Loot".
4. Unsigned, Blizzard block beginning "The latter".
5. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "Announcement at the ECTS".
6. Staats, *The WoW Diary*, section "Gameplay".
7. Staats, *The WoW Diary*, section "September 2002: Internal Alpha 1.0".
8. Staats, *The WoW Diary*, section "October 2002: Still Unanswered Questions".
9. Staats, *The WoW Diary*, section "September 2002: Internal Alpha 1.0".
10. Staats, *The WoW Diary*, section "September 2002: Internal Alpha 1.0".
11. Garrett Hardin. The Tragedy of the Commons. *Science*, 162 (3859):1243–1248, 1968.
12. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
13. Staats, *The WoW Diary*, section "July 2004: Public Beta 3.0".
14. To gank is to kill a player who has no real chance of fighting back. Binding sets the place a character can return to.
15. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
16. Staats, *The WoW Diary*, section "December 2002: Blizzard Looks to Asia".
17. Hazzikostas, second answer.
18. Blizzard Entertainment. Raid Finder Q&A. `worldofwarcraft.blizzard.com`, n.d., answers 6 and 8.
19. Transmog, short for transmogrification, lets a player give one item the appearance of another, so old items are collected for their looks.
20. Unsigned. Personal Loot and Legacy Loot Mode in Battle for Azeroth Pre-Expansion Patch (embedded Blizzard reply). Wowhead News (post by Squishei), 2018, embedded official post.
21. Unsigned, Wowhead's text around the embedded post.
22. Unsigned, Blizzard block beginning "The latter".
23. Mark Rosewater. Eternal Formats of the Spotless Mind. Making Magic, `magic.wizards.com`, 2016, section "A Long Time Ago".
24. Rosewater, paragraph beginning "Which brings us to 1995".
25. Rosewater, section "A Long Time Ago".
26. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 23 at 1:27.
27. Wizards of the Coast. Magic: The Gathering Tournament Rules (effective February 27, 2026). `media.wizards.com`, 2026, section 5.5.
28. Mark Rosewater. Twenty Years, Twenty Lessons, Part 2. Making Magic, `magic.wizards.com`, 2016, section "Lesson 9".
29. Distraction Makers, episode 23 at 4:49.
30. Distraction Makers, episode 10 at 14:09, 25:57 and 26:14.
31. Distraction Makers, episode 23 at 4:49, 6:25 and 6:49; episode 16 at 22:09.
32. Distraction Makers, episode 16 at 0:52, 1:20 and 9:49.
33. Distraction Makers, episode 16 at 2:13 and 20:46.
34. Distraction Makers, episode 22, part B, at 5:17, 13:08 and 13:53.

35. Thomas C. Schelling. *The Strategy of Conflict.* Harvard University Press, 1960.
36. Mark Wallace. Report of Rob Pardo's keynote on World of Warcraft at the Austin Game Conference. Game Developer, 2006, paragraph 10.
37. Staats, *The WoW Diary*, section "March 2001: My First Six Months".

Table 17.1. Seven behaviors and where each game put them. "Not stated" means the sources read for this book are silent; "No counterpart" and "The object of the game" mark how the behavior maps onto that game.

Behavior	*World of Warcraft*	*Magic*
Who takes a dropped item	Open in the first alpha, no norm recorded; a leader under master loot; code under personal loot; a roll from 2022	No counterpart
Harm done through the world (the campfire)	Code: the designer concluded player campfires should do no damage (diary, September 2002)	Not stated
A trade other than the one agreed	Code: a confirmation button (diary, "Gameplay")	Not stated
Killing other players	The player's choice of server (diary, June 2003 to July 2004)	The object of the game; no counterpart
Pace of play	Not stated	Written rule with a judge (tournament rules, section 5.5)
Copies of one card in a deck	No counterpart	No limit in 1993; four, in written rules from 1994 (Rosewater)
How strong a deck to bring to a casual table	No counterpart	Norm

Part V

Economies, Rewards and Progression

Chapter 18

Loot Is an Economy

A designer is adding swords. Monsters in this game drop items at random, and the patch on the designer's desk holds three new ones: a saber, a cutlass and a scimitar. They are ordinary weapons, there to give the common tier some variety, so each goes into the loot table at the same weight as the common swords already listed. The work takes a few minutes and looks like no decision at all.

The case is a teaching example from Candlesan, who designed monsters for *Diablo III* and raid bosses for *World of Warcraft*, written for anyone building an item-based game with random loot.[1] The three swords diluted the pool and lowered the drop rate of every epic and legendary item.[2] Nobody opened the legendary's row. The best item in the game became harder to find because the list got longer.

The coupling runs the other way too. With the game live, a bug turns up in the scimitar, and a hotfix disables the drop until it can be repaired.[3] With one entry gone, every other item drops more often, legendaries included, and players get used to the higher rate. Then the bug is fixed, the scimitar comes back, and a team that has just repaired a sword has nerfed the legendary drop rate without meaning to.[4]

Three changes, none of them about legendaries. The legendary rate moved all three times, down, up, then down again. The last move is the one players feel, and nobody at the studio chose it.

18.1 A Price Nobody Typed

The ***drop rate*** of an item is how often a kill produces it, and from the player's chair it is a price. An item that drops once in fifty kills costs, on average, fifty kills, paid in evenings. Whoever sets the rate sets what the best thing in the game costs and how many copies of it exist. How loot feels is Chapter 6's subject and how strong an item is belongs to Chapter 19. This chapter is about control. The legendary rate is the number players care

about most, and the table around it is edited for years for other reasons, a rarity revised or junk added to be salvaged for crafting.[5] The rate has to survive all of it and change only when someone means it to.

The first table most developers write puts every item in one list and rolls against it. Rarity is handled by repetition: a common gets six entries, an epic two, the legendary one.[6] Literal entries need renumbering whenever something is added, so most teams give each item a weight and let the code work out the ranges.[7] Adding a sword becomes a single line. The design works in development and, for a simple game, is probably enough to ship.[8] Weights of 6, 3, 2, 1 look like permanent policy.

The arithmetic below uses one such table: weight 6 for each common entry, four rares at weight 3, two epics at weight 2 and one legendary at weight 1. The rare tier and the item counts are added to the six, two and one of the example.

A weight is a raffle ticket, and the chance of winning depends on how many tickets everyone else holds. The legendary's chance on a kill is its weight divided by the sum of every weight in the table. The numerator is the legendary's. The denominator belongs to the whole item list, and every row added anywhere makes it larger.

With five common entries the weights sum to 47 and the legendary drops on 2.128 percent of kills. With ten common entries the rate is 1.299 percent. With forty it is 0.389 percent. Figure 18.1 plots the slide against a roll that names the legendary rate outright and stays at 2.0 percent however long the list grows.

Turn each percentage into the average number of kills before the first legendary. At five commons it is 47 kills. At ten it is 77. At forty it is 257, more than five times the wait at five, for the same item from the same monster. The legendary holds a single ticket, so the average wait equals the total number of tickets, and each new common adds 6 kills to it. The three swords cost every player 18 kills per legendary. Run it backward for the hotfix: a table trimmed from fifteen commons to ten sees its legendary rate rise from 0.935 percent to 1.299, more than a third higher, and the wait fall from 107 kills to 77.

The failure is organizational before it is arithmetical. The person adding swords is making a content decision. Nothing in the task says supply, and nothing in the file shows the rate. Under weights the rate is never typed anywhere: it is what is left after everyone has finished typing. How many legendaries exist in the world is decided, release after release, by whoever touched the table last, and that person does not know it.

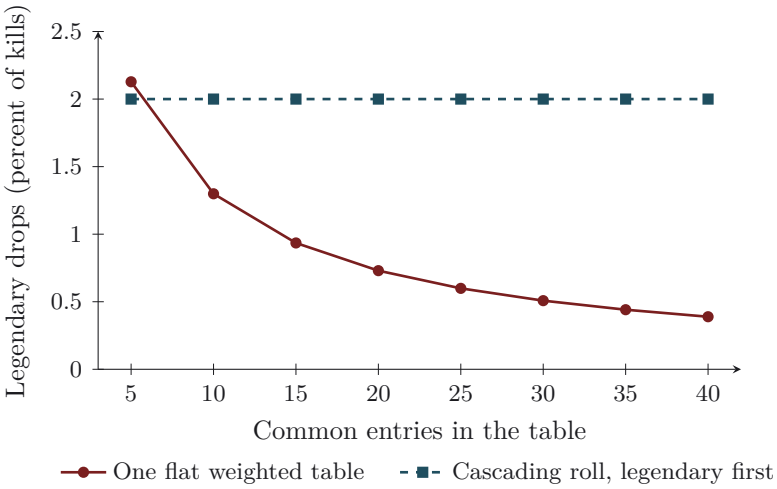


Figure 18.1. The legendary drop rate as common entries are added. In the flat table each common entry has weight 6, four rares have weight 3, two epics weight 2 and one legendary weight 1. The weights are illustrative, patterned on the example in Candlesan's commentary, and describe no shipped game. The cascading roll is held at 2 percent.

18.2 Players Forecast What Falls

Players never see the table. They see what drops and calibrate to it, which is why the scimitar's repair arrives as a loss. They also forecast. A 2017 Blizzard forum post on the legendary items of the *Legion* expansion reported that players were predicting their odds from historical data, and that the data was badly out of date because the chances had since been raised.[9] Players forecast from experience, so a rate that drifts makes the forecast wrong without telling anyone why.

To the player, a rate changed on purpose and a rate changed by a longer sword list are the same event. The flat table manufactures balance changes nobody announces, because nobody at the studio knows there is anything to announce. Of the sources of uncertainty Chapter 5 draws from Costikyan, two meet here.[10] Randomness is the roll, the uncertainty a loot system exists to deliver. Development anticipation is not knowing what the next patch will do. The flat table converts the second into the first in silence. A player facing it is uncertain about the roll and about the dice.

Second-Order Effect: The price of good luck

A rate can be tied to something worse than a list of swords. At launch, *Legion*'s legendaries ran on a rule under which each legendary a player earned lowered the starting chance of the next. The team's long-term goal was for players to own several and swap them as the situation required, and the designers did not want past good fortune to mean harder times ahead. So the chances were raised more than once, and every legendary after a player's second was given the same chance, the third costing the same luck as the eighth.[11] The launch rule put a rising price on success, working against the purpose the items served, and the repair is what left players forecasting from data that no longer applied. Chapter 20 takes up what players remember of rates like these.

18.3 Decide How Rare Before Deciding What

The repair comes in two steps. The first divides the items into separate tables by rarity, four in the example, and rolls the tier up front; a second roll picks an item inside that tier's subtable.[12] Three new common swords now compete only with other commons. The second step is the ***cascading roll***. When an item is about to drop, roll for the legendary first, at a stated chance (2 percent in the example). Only if that misses, roll for the next tier down, and so on until a tier hits; then pull from that tier's table. What it buys is precision and stability during live operations (Figure 18.2).[13]

In Figure 18.1 the cascade is the line that does not move: 2.0 percent with five commons and with forty, an average wait of 50 kills in both. It is not more generous. At five commons the flat table pays better, 2.128 percent against 2.0, 47 kills against 50. The cascade gives the designer a number that is typed and can be found.

A probability that players track should be a parameter someone sets, not the remainder of a sum.

The dilution has not disappeared. A new common still lowers the odds of every other common, so the cost of an edit lands inside the tier where it was made. And dependence now runs one way. Change the legendary's chance and every tier beneath shifts a little, since they share what the legendary roll leaves. Change anything beneath, and the legendary does not move.

In terms of ***expected value***, the average wait for an item is one divided by its rate, and under a flat table the price of the most valuable good is a

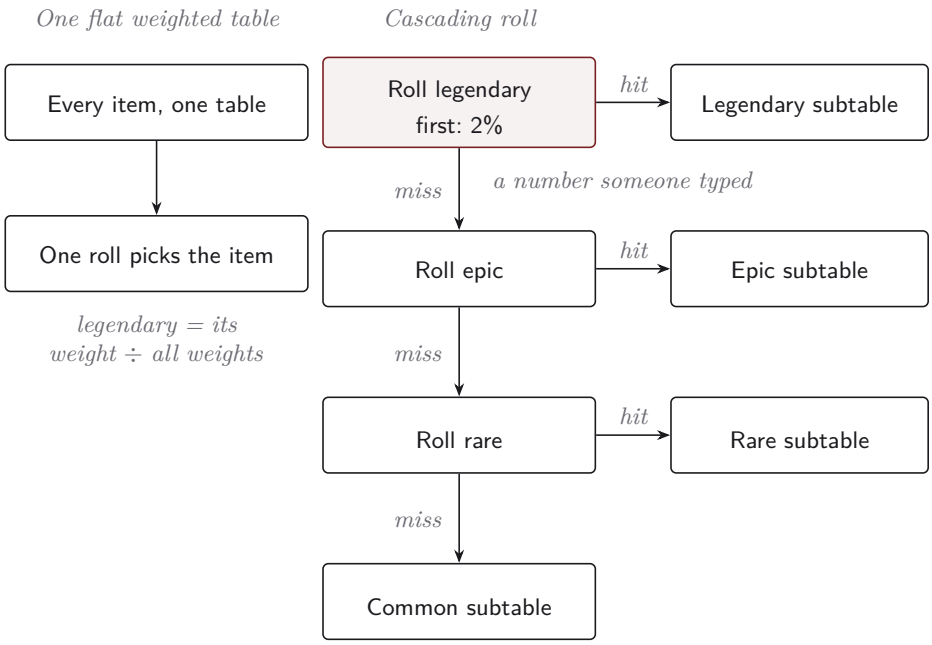


Figure 18.2. The two loot tables of Candlesan's teaching case. In the flat table the legendary's chance is its share of every weight in the list; in the cascade it is the first roll, at a number someone typed. The four tiers follow the illustrative table of Figure 18.1.

function of every other row. A quantity has ***invariance*** to an edit if the edit cannot change it.

Design equation (a conceptual framework, not a law)

flat table: chance of the legendary = its weight ÷ the sum of all weights

cascading roll: chance of the legendary = a number someone typed

The first line is a share of a total, and a share moves whenever the total does. The second is invariant to everything beneath it. What the hotfix did to the flat table is the game's version of what economics calls a supply shock: a change in how much of a good is produced, arriving from outside the market for that good. A disabled scimitar is a strange origin for one.

18.4 More Rows Than Slots

A live table is pushed in one direction, toward more rows. In October 2003, by John Staats's diary of *World of Warcraft*'s development, monster loot felt too repetitive and limited, so the game borrowed the semi-random item generation of *Diablo* and moved better items to vendors because players were not using merchants enough: "there was evidence that we needed more rewards and bigger loot tables." Later that month procedurally created items were added to keep the tables fresh.[14] The beta of August 2004 brought a loot system that reflected monster level,[15] and Chapter 19 follows how often the stats on all those items were rewritten.

The diary also states the constraints as a system. Players measure progress by the equipment they acquire. Gear cannot differ so much that casual players cannot compete, yet high-end rewards have to justify their extra time, a balance Chapter 14 follows into one raid's difficulty tiers. The rewards of one activity cannot supplant another's, and there are far more activities and zones than item slots.[16] That is allocation under scarcity.

Supply is half of an economy; the other half is who ends up holding the goods. Chapter 1 tells the alpha in which anyone could loot any corpse. A flawless table can still fail at delivery, and among the small decisions credited with stabilizing the game economy are no-drop rares and confirmation buttons on trades,[17] the same rules Chapter 17 reads for their first purpose, preventing griefing.

The team came to see items as an economy on paper in March 2002. After the Game Developers Conference, Eric Dodds sent the team notes on a talk by Richard Garfield about games of tradable objects, drawing heavily on *Diablo* and *Magic: The Gathering*, and admitted he had not thought of their game as exactly that kind of game. Three of the terms concern items. A vanity object looks different and plays the same. A bad object is clearly worse than others of its type, which lets players feel they chose well and raises the value of the good ones. Item retirement is the need for old items to leave if new ones are to excite anybody, by finite charges (which the notes dismiss), by a finite lifespan as with tournament *Magic* cards, or by new items simply being better. The advice: many vanity items, some deliberately weak items at each level, and items that are no-drop or wear out so slowly the player hardly notices, because "the item should leave the world eventually."[18]

Garfield's own game sets its drop table in public. A booster pack of *Magic* is a loot drop bought in a shop, and the Play Booster, announced in 2023 and first sold with *Murders at Karlov Manor* in February 2024,[19] runs slot by slot: six commons; a seventh slot that is a common seven times in eight and otherwise a card from a special list; three uncommons; one rare or mythic rare, a rare six times in seven; then a land, two wildcards and a card not for play, with the percentages those of one named set.[20] The pack is built rarity first: each position's tier is fixed, or rolled at a stated chance, before any card is chosen, the first of the two steps made of cardboard. Reworking it for drafting shifted the ratio toward commons but kept the chance of opening several rares or mythics, the reason players most often gave for liking the earlier product.[21]

Magic states the bad object as policy: power is relative to the surrounding cards, so not every card can be good, and good cards that look bad cannot be made without also making cards that are bad.[22] Chapter 16 reads that column for the complaint behind it. Scarcity of quality is what lets a find be a find. Rarity also rations exposure to ideas, which Chapter 21 follows.

The rate also paces play. Pieces that reach players slowly and at random carry variance into collecting: nobody holds everything at release, so the population takes longer to solve the environment.[23] A drop rate sets how soon a population is fully equipped, and so how soon it runs out of things to look for.

18.5 Where a Fixed Rate Stops Protecting

The cascade fixes the rate of a tier, not of an item. Add a second legendary to the legendary subtable and the tier still hits at the stated rate while the chance of a particular legendary is divided (Table 18.1).

Legendaries	Tier rate (%)	One named item (%)	Average kills for it
1	2.0	2.00	50
2	2.0	1.00	100
3	2.0	0.67	150
4	2.0	0.50	200

Table 18.1. A cascading roll held at 2 percent as legendaries are added to its subtable. The first column counts the legendaries in the subtable. The tier rate holds and the rate of any one named item divides. The numbers are illustrative.

With one legendary the named item costs 50 kills on average. With two it costs 100, and with four it costs 200, at a tier rate that never left 2.0 percent. A player hunting one weapon for one class is back in the opening, one level down. Invariance has to be bought again at every level where players care, and the designer has to know which level that is.

A fixed rate is also a promise about an average, and no player lives the average; Chapter 20 follows the capped currency with which a later season went around the table altogether. And some rates should move, by decision. The same chapter follows the *Cataclysm* legendary that was to be made rarer on purpose, and Azerite armor, built to take back with every upgrade until the rule was withdrawn. A parameter nobody may ever touch is a second way of losing control of it. The cascade itself is a primer, and a complex game adds conditionals and nested tables on top of it,[24] each one a new place for a rate to depend on something nobody is watching.

The Designer's Test

1. Which probability in this system do players track, and is it a number someone typed or the remainder of a sum?
2. What unrelated edit would change it, and who on the team makes that kind of edit?
3. If one drop were disabled for a week, which rates would move, and would players notice when they moved back?

4. At which level do players care: the tier, or one named item inside it?

5. How does an item leave the world, and is there a rate here that should change over time by someone's decision?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V1 at 0:04; V5 at 0:24 and 1:59.
2. Candlesan, commentary, V1 at 0:50.
3. A hotfix is a small change applied to a running game without waiting for a full patch. To nerf something is to make it weaker.
4. Candlesan, commentary, V1 at 1:03 and 1:13.
5. Candlesan, commentary, V1 at 1:45.
6. Candlesan, commentary, V1 at 0:07.
7. Candlesan, commentary, V1 at 0:31.
8. Candlesan, commentary, V1 at 0:42.
9. Ornyx. Legendary Drop Rates. World of Warcraft Forums, `us.battle.net` (read from Wayback copy), 2017.
10. Greg Costikyan. *Uncertainty in Games*. The MIT Press (Playful Thinking series), 2015.
11. Ornyx.
12. Candlesan, commentary, V1 at 1:26.
13. Candlesan, commentary, V1 at 2:05.
14. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "October 2003: Free Pizza and Other Hardships".
15. Staats, *The WoW Diary*, section "August 2004: Public Beta 4.0".
16. Staats, *The WoW Diary*, section "Why MMOs Are So Difficult to Create".
17. Staats, *The WoW Diary*, section "Gameplay".
18. Staats, *The WoW Diary*, section "March 2002: Competitive Collaboration".
19. Wizards of the Coast. Collecting Murders at Karlov Manor. `magic.wizards.com`, 2024.
20. Mark Rosewater. What Are Play Boosters? *Making Magic, `magic.wizards.com`*, 2023.
21. Rosewater.
22. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, Sections 1 and 6.
23. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 7 at 3:20, 4:13, 5:22 and 6:00; episode 23 at 12:09 and 12:40.
24. Candlesan, commentary, V1 at 2:27.

Chapter 19

Progression Is a Mathematical Promise

Armor could hardly be simpler than this. The enemy attacks for 50. You are wearing 10 armor. You take 40. Candlesan, the designer whose commentary supplied Loatheb in the introduction, says that when work began on *Diablo II* the team took that simple route, the subtraction *StarCraft* uses.[1] It failed at once.

Take two monsters. The first is slow and hits for 100 damage every 2 seconds. The second is fast and hits for 10 damage five times a second. Both deal 50 damage per second, so by the usual measure they are the same threat.[2] Put 10 points of subtracting armor on the character they attack. Against the slow monster the armor does almost nothing. The fast monster's damage falls to zero.[3]

Table 19.1 runs the two monsters through four amounts of armor. With 10 armor the slow hitter keeps 45 of its 50 damage per second, a loss of 10 percent, and the fast hitter keeps none. Halve the armor to 5 and the gap is still wide: the slow hitter loses 5 percent and the fast one 50 percent, ten times the share.

Flat armor	Slow, DPS kept	Fast, DPS kept	Slow, lost	Fast, lost
0	50.0	50.0	0.0%	0.0%
5	47.5	25.0	5.0%	50.0%
10	45.0	0.0	10.0%	100.0%
20	40.0	0.0	20.0%	100.0%

Table 19.1. Two attackers with the same raw 50 damage per second against armor that subtracts a fixed amount from every hit. The slow hitter deals 100 every 2 seconds and the fast one deals 10 five times a second, the values in Candlesan's example (V7 at 0:41). Author's calculation.

A player who equipped those 10 points was told something by the game: you are now harder to kill. How much harder depends on a fact the player never chose and may never see, how the monster's damage happens to be sliced. Against one enemy the purchase is nearly worthless. Against the other it is immunity.

19.1 Subtraction and Percentages Both Break

A game with progression and hundreds of unit types cannot: it needs attack speed to be an aesthetic choice and not a balancing point.[4] A wasp's fast sting and an ogre's slow club should not rewrite the value of every helmet. Attack speed is only the first shape damage comes in. A poison that deals 500 damage over 15 seconds, fire on the ground and reflected damage would each need a special rule under subtraction.[5]

A stat on an item is a promise, and under flat armor the designer cannot say what was promised. Ten armor is worth everything against small hits and less and less as monsters grow and their hits grow with them. Progression over hundreds of hours needs each point to mean something stated in advance.

The obvious repair takes hit size out of the rule: each point of armor is 1 percent less damage, so 10 armor is 10 percent and 50 armor is 50 percent.[6] A share of every hit treats the two monsters, the poison and the fire alike, and 50 on the character sheet means half. It is the naive approach, and its flaw cannot be seen from a single hit.[7]

Armor now gets more valuable the more of it you have. Going from zero to 10 percent is worth very little. Going from 80 to 90 percent halves the damage that gets through.[8] The same ten points on the tooltip double how long the character lives.

The quantity that shows this is ***effective health***: the raw damage a character can absorb before dying, which is health divided by the share of damage that gets through. Take an illustrative character with 10,000 health, under a linear rule in which 20,000 armor would be 100 percent. With no armor, effective health is 10,000. The first 2,000 armor, zero to 10 percent, raises it to 11,111, a gain of 1,111. The same 2,000 armor added between 16,000 and 18,000, from 80 to 90 percent, raises it from 50,000 to 100,000, a gain of 50,000 and 45 times as much for the same purchase. At 19,000 armor effective health is 200,000. At 20,000 it has no finite value.

With health H, armor a and A the armor that would reach 100 percent, effective health under the linear rule is $H/(1 - a/A)$, and one more point

adds $HA/(A-a)^2$. The denominator shrinks as armor rises, so each point is worth more than the last, without limit. In play, the only armor that matters is the last of it.

So games that take this route end up needing a hard cap on armor.[9] The cap is a patch, and players answer it the only sensible way. If every point is worth more than the one before, buy all of it. The cap becomes the build: below it every piece of gear without armor is an error, and at it every further point is worth nothing. The stat has become a threshold to reach and then ignore.

19.2 Make Survival the Straight Line

The formula *World of Warcraft* shipped for armor gets the opposite complaint. People say it has diminishing returns,[10] and they are reading a real column. In the illustrative model the first 5,000 armor buys 50.0 points of damage reduction, the next 5,000 buys 16.7, the 5,000 after that 8.3, and the next 5,000 buys 5.0. Armor looks stale.

That is the wrong way to look at it.[11] Count hits survived instead. Take 1,000 life and an enemy that hits an unarmored character for 100: ten hits. Add armor 100 points at a time. The percentage slows. The hits survived go 10, 20, 30. "Every 100 points of armor buys exactly 10 more hits."[12] Survivability is directly proportional to armor.[13]

Decide which quantity the player should feel growing in a straight line, then derive the displayed number from it.

Work backward from the promise. Ask for effective health that rises by the same amount with every point of armor, and exactly one mapping from armor to a percentage delivers it. Unlike the book's other boxes, this one is exact arithmetic inside the model, where every hit is reduced by the same share.

The mathematics

$$DR = \frac{a}{a+K} \qquad EH = \frac{H}{1-DR} = H\left(1+\frac{a}{K}\right)$$

a is armor, H is health, DR is the share of damage removed, EH is effective health, and K is a constant the designer chooses. One more point of armor adds H/K to effective health, whatever a already is.

K is the armor at which half the damage is removed. In the illustrative model K is 5,000 and health 10,000, so every point of armor adds exactly 2 to effective health, the first point and the twenty-thousandth alike. Against hits of 1,000 the character survives 10 hits with no armor, 20 at 5,000 armor, 30 at 10,000, 40 at 15,000 and 50 at 20,000, while the percentage reads 0, 50.0, 66.7, 75.0 and 80.0. In Candlesan's example the first 100 armor doubles the hits survived, so its constant is 100. This is the equation of *Diablo II*'s defense rating and of *World of Warcraft*'s armor;[14] the K and H here are the script's.

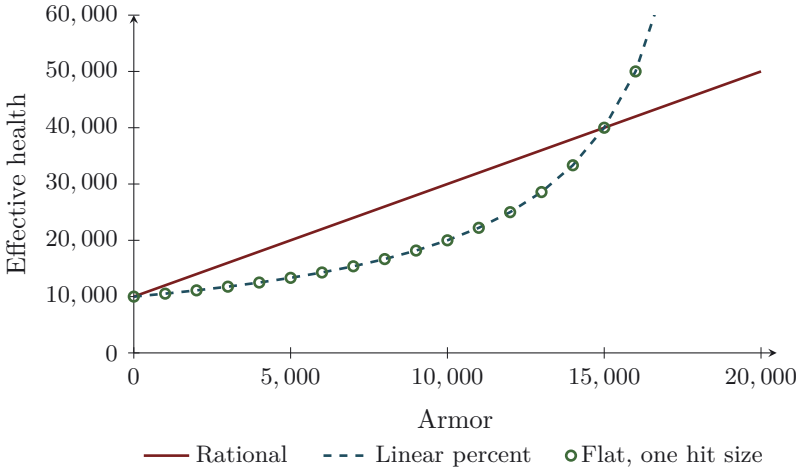


Figure 19.1. Effective health of an illustrative character with 10,000 health under three armor rules. The rational rule ($K = 5{,}000$) adds 2,000 for every 1,000 armor. The linear rule (100 percent at 20,000 armor) lags it, meets it at 15,000 armor, where both stand at 40,000, then leaves the chart: 200,000 at 19,000 armor and no finite value at 20,000. The flat rule is priced here against a single hit of 1,000, so its points fall on the linear curve. Its failure is in Table 19.1, not in this picture.

Figure 19.1 puts the three rules on the axis that matters. The linear rule looks orderly on a tooltip and bends upward out of sight. The rational rule, so called because its percentage is armor over armor plus a constant, flattens in percent and runs straight in survival. It still takes a share of every hit, so slow and fast attackers stay equal, and it needs no cap: the percentage approaches 100 without arriving, 80.0 percent at the chart's right edge, where the linear rule had already reached invulnerability.

What separates the three rules is ***marginal value***, what one more unit buys given what is already owned. It depends on hit size under flat armor, rises to the cap under linear percent, and is constant, H/K, under the ratio-

nal rule. Behind that is the choice of ***state variable***, the quantity a model keeps well behaved. Keep the percentage linear and effective health bends toward infinity; keep effective health linear and the percentage flattens. The designer chooses which one gets to be simple, and the other absorbs the curvature.

19.3 A Combat Formula Is a Price List

Every item in a game is priced by its combat numbers. Once *World of Warcraft*'s combat numbers were real (Chapter 22 tells how long they were a mock-up), items were especially hard to make during the first internal alpha, because their values had to be changed by hand every time combat was tweaked, almost daily.[15] By the second alpha there were fifteen hundred items.[16] Two years later, in the final month, the designers "changed the stats on every item (again)."[17] Each change to the formula re-prices everything sold under it.

The same numbers set the pace of play. In the first alpha Allen Adham cut monster hit points by 33 percent to speed up leveling, and by the second, monster health, armor and damage had all come down by a third for the same reason.[18] In mid-2003 Kevin Jordan argued that fights left too much downtime and felt too close to an older online game, where one monster of a character's own level took most characters below half their health. The plan was to lower monster damage so monsters could have pets and travel in pairs, and an accidental extra pull would not mean death. The team expected that small cut to move the pace of combat, grouping and the appeal of playing alone.[19] How much a character absorbs, set against how fast enemies deal damage, is the number of mistakes a fight forgives. Choosing the quantity that grows in a straight line fixes how much more forgiveness each upgrade buys.

Progression is also kept in a second currency. When the August 2004 beta revised combat and loot together, "the item stats had bigger numbers, so players perceived them to be more powerful even though they were not."[20] Rob Pardo's philosophy pointed the same way: every unit and class should feel as though it cannot be stopped.[21] Power as the formula counts it and power as the player perceives it are separate ledgers, and the second can move while the first stands still.

The rational formula pays for the same gap. The percentage is on the character sheet and hits survived is on no screen, so by the ***salience*** of Chapter 1 players quote the column that bends and call the stat weak. A

promise players are meant to feel has to be shown in the quantity that keeps it.

Second-Order Effect: Numbers too big to read

A later Blizzard design post said health, damage and other stats had become too large for players to grasp, and that older tiers of content left sudden spikes in power. The repair shrank the whole scale and kept every ratio: a Fireball that hit for 450,000 against 3,000,000 health would now hit for 30,000 against 200,000, 15 percent both times. The post then had to tell players this was not a nerf.[22] A 2020 post did the same to character levels, turning level 120 into level 50, fewer than half as many, so that each new level would feel more meaningful and leveling would be faster.[23] The rational formula keeps one stat honest, but each point of armor adds H/K, so raising health makes every point worth more, and a character who gains both for years grows faster than either line. A scale with no cap and no top is eventually rescaled. A player who watched a Fireball land for 450,000 measures the new 30,000 against the old figure, and outcomes judged as gains and losses from a reference point are what Kahneman and Tversky's prospect theory describes.[24] A post that says the ratio is unchanged is arguing with a memory. Chapter 25 returns to the squish.

19.4 Mana Is the Denominator

Magic: The Gathering has no armor stat. It has the quantity everything else is priced against, printed in the corner of every card: a card is worth what it does for the mana it costs. Mark Rosewater's checklist of what a set needs treats costs as a distribution, something to play at each point of the scale, checked at each rarity, and asks for mana sinks, effects that give players something to spend on late in the game, when they have more mana than cards.[25] A sink keeps the tenth mana worth something, as the rational formula keeps the ten-thousandth point of armor worth something.

Mana comes in whole steps. No color gets a vanilla 2/2 for one mana, because Savannah Lions marks the ceiling R&D wants at that cost, and the same 2/2 for two mana, Grizzly Bears, is not strong enough for tournaments. Nothing can be costed between the two.[26][27] Armor can be tuned by a single point; a creature's cost cannot.

A second currency is a formula's worst case. Chapter 23 follows energy to *Kaladesh*. Its team knew energy would need an economy, built a framework in which equal energy costs bought roughly equal power, and limited how often energy and mana appeared together in a cost, so that energy would not read as a substitute for mana.[28] The January 2018 announcement that banned Attune with Aether and Rogue Refiner described them as cards producing incidental, free energy beside an effect almost strong enough to play without it.[29] A currency meant to have an economy was being handed out at a price of zero, on cards already worth their mana. Chapter 25 tells the bans. Investigate shows the milder repair: it first drew a card outright and was changed to make a Clue that costs mana to cash in.[30] The reward stayed and was run back through mana.

A card game makes its progression promise to the buyer of the next set. Would anyone buy the next small expansion if none of its cards were good enough for tournaments?[31] Keeping that promise by the linear route has a name. The danger of power creep is that sets must keep getting stronger to compete with the sets before them, because power is relative, while fun is a threshold: one thing can be as fun as another without beating it. The stated policy was to make every set as fun as possible and not to compete with earlier sets on power.[32] *World of Warcraft* could shrink its numbers and keep the ratios. A card already printed keeps the numbers it was printed with.

19.5 When a Cap Is the Right Answer

The objection to the capped percentage holds for a game that wants armor to keep increasing.[33] A game with a short arc, or one whose designers want a stat to be finished at some point, can take the linear percentage, state the cap and plan for players to reach it. The failure in this chapter is a cap added late to stop a curve, not a cap chosen in advance as the place the stat ends.

The Designer's Test

1. What does one more point of this stat buy at the bottom of the scale, and what does it buy at the top?
2. Which quantity does the player feel during play, and is that the one you have made linear?

3. Does the value of the stat depend on something meant to be cosmetic, such as how an enemy's damage is divided into hits?
4. If the stat needs a cap, what will players do below it, at it, and above it?
5. Which number does the player see, and what will this scale look like after ten more tiers?

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V7 at 0:19 to 0:32.
2. Damage per second, usually shortened to DPS, is total damage divided by time. It is the standard yardstick for comparing attackers that hit at different speeds.
3. Candlesan, commentary, V7 at 0:37 to 0:59.
4. Candlesan, commentary, V7 at 1:01.
5. Candlesan, commentary, V7 at 1:12.
6. Candlesan, commentary, V7 at 1:27.
7. Candlesan, commentary, V7 at 1:32.
8. Candlesan, commentary, V7 at 1:43.
9. Candlesan, commentary, V7 at 1:56.
10. Candlesan, commentary, V7 at 2:13.
11. Candlesan, commentary, V7 at 2:15.
12. Candlesan, commentary, V7 at 2:17.
13. Candlesan, commentary, V7 at 2:50.
14. Candlesan, commentary, V7 at 2:05.
15. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "September 2002: Internal Alpha 1.0".
16. Staats, *The WoW Diary*, section "November 2002: Internal Alpha 2.0".
17. Staats, *The WoW Diary*, section "September 2004: Going Gold".
18. Staats, *The WoW Diary*, sections "September 2002: Internal Alpha 1.0" and "November 2002: Internal Alpha 2.0".
19. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
20. Staats, *The WoW Diary*, section "August 2004: Public Beta 4.0".
21. Frank Cifaldi. Report of Rob Pardo's talk Making a Standard: Blizzard Design Philosophies, Game Developers Conference. Game Developer, 2010, section "Make Everything Overpowered".
22. Blizzard Entertainment. Dev Watercooler: Pruning the Garden of War. `worldofwarcraft.blizzard.com`, n.d., section "Stat Squish".
23. Blizzard Entertainment. Explore the Updates Arriving with the Shadowlands Pre-Expansion Patch. *Blizzard News,* `news.blizzard.com`, 2020, section "Levels, Items, and Stats Adjusted".
24. Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, 1979.
25. Mark Rosewater. Nuts & Bolts 15: Structural Support. Making Magic, `magic.wizards.com`, 2023, sections "Mana Curves" and "Mana Sinks".

26. Mark Rosewater. Let's Start at the Very Beginning. *Making Magic,* `magic.wizards.com`, 2003, section "Grizzly Bears".
27. A vanilla creature has no rules text, only its two numbers: power, the damage it deals, and toughness, the damage it takes to destroy.
28. Mark Rosewater. Kaladesh Ingredients, Part 1. Making Magic, `magic.wizards.com`, 2016, section "Energy Conservation".
29. Ian Duke. January 15, 2018 Banned and Restricted Announcement. *`magic.wizards.com` (read from Wayback copy)*, 2018, section "Temur Energy and Other Energy Variants".
30. Mark Rosewater. Looking Back, Part 2. Making Magic, `magic.wizards.com`, 2024, section "2016 Best Addition".
31. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, Section 1.
32. Mark Rosewater. In the Mailbag. *Making Magic,* `magic.wizards.com`, 2008, reply beginning "You caught us".
33. Candlesan, commentary, V7 at 2:01.

Chapter 20

The Economy Remembers

Beta players of *World of Warcraft* met a rule under which a character who stayed in the world long enough went on half pay. Experience, what a character earns from kills and quests on the way to the next level, dropped from 100 percent to 50 percent after a couple of hours of play, the figure Rob Pardo gave at GDC 2010. Beta players hated it, universally.[1]

The rule was written to take nothing from anyone. At a team-wide design meeting in October 2003, John Staats's diary records, the designers set out to limit how long players could grind through levels, because they did not want a game in which people lost contact with the real world. Lowering the experience earned after six or so hours of play would encourage a break.[2] The talk says a couple of hours and the diary says six or so. They agree on the shape: a fresh character earned the full rate, a tired one earned less, and an intent of care was received as a fine.

The fix changed only the framing. Players would now start at 200 percent and fall to 100 percent, the identical mechanic presented as a bonus where it had been a punishment.[3] Under either version the rested character earns exactly twice what the unrested one does.

Chapter 1 met this pair of outcomes as a limit on its own argument: a rule's incentive predicts what players do, and it does not predict how they feel. This chapter starts there. No table in the game held what a player used to earn. The players responded as though something did.

20.1 Players Measure from Where They Stand

A designer has to move rewards both ways. Rates have to come down when a mistake made something too generous, when one class runs away with the game, or when a new tier needs room above the old one. A live game that can only raise its numbers has a ratchet in it, and every reduction is made in front of a population already adapted to the current level.

The natural review treats a reward system as a schedule of payoffs. If the new schedule pays what the old one paid, nobody loses; if a rate drifted above its intended value by accident, returning it restores the design. On that view the two rest designs are the same curve with the labels moved. Most reward changes are still reviewed by comparing totals.

Nobody experiences a schedule as a schedule. A player meets it in order, one hour after another, and judges each hour against the one before. In the first version the session opens at 100 percent, the normal rate and the best available, and some hours in it becomes 50: the player has been moved below the only standard the game ever offered, for having played. In the second the session opens at 200 and ends at 100. The fall is the same size in proportion and ends on what reads as ordinary pay.

Players do not carry the designer's baseline. They carry a ***reference point***, the level they last got used to or were told to expect, and score each change as a gain or a loss from there. The first design put the reference at the top of the schedule and showed the drop; the second put it at the bottom and showed the surplus. A review of totals cannot see the difference, because the difference is not in the totals.

A reward system has a memory, and the memory is in the players: every change is measured from the level they are used to, whatever level the designer intended.

Three consequences follow. Generosity is cheap to grant and expensive to withdraw, because the grant is scored as a gain from the old reference and the withdrawal as a loss from the new one. An accident counts as much as a decision, because players adapt to what they receive, not to what was meant. And the first level a system ships at, or is shown at, is a commitment: it sets the point from which every later correction will be judged.

The attachment forms fast. In the fourth phase of the public beta, after a round of class balance changes drew the usual negative reaction, the diary states the rule in passing: "Players with overpowered skills and items grew to love them and vented whenever they were 'nerfed.'"[4] These were testers of an unfinished game, and they owned their provisional powers anyway. Players acclimate just as quickly to a better rate nobody announced, like the hotfix in Chapter 18,[5] and they keep the books themselves, forecasting their legendary odds from history the code never stored.[6]

20.2 Too Strong, Too Sure

The death knight was a new class in a game whose other classes players already knew, and the designers feared a brand-new class that felt neglected, so they overdid it. Bringing it back into line left its players feeling that the rug had been pulled out from under them, and the swings cost the team some of the players' trust; the plan for the monk was a simpler resource model and fewer balance swings.[7] A generous start set the reference, and a correction that only returned the class to where it belonged was read as confiscation. A large swing also teaches players not to rely on the numbers under them, a lesson that outlives the patch. The fear was sound: a class that launches too low fails on its first day in front of the players it most needed to please; starting low trades a later nerf for an earlier disappointment.

The *Cataclysm* legendary is the opposite error with the same root. Nothing about it was too strong. It was too sure: a raiding guild could complete it with certainty, many specializations could use it, and guilds came to expect legendaries in turn. Greg Street's post-mortem judged that it had become an entitlement and stopped feeling like a reward, a bar to be filled, and the remedy was to make the next legendaries rarer, so that not every raiding guild would obtain one.[8]

> **Second-Order Effect: Rarer, and unfinished**
>
> The same paragraph names the bill: under a rarer design, some guilds will almost complete a legendary.[9] A certain reward produced entitlement. An uncertain one produces groups who invested the effort and hold nothing, and the effort already spent is exactly what makes stopping hurt. The fix exchanges one grievance for another and bets that the second is the healthier one.

The studio's later record works around the asymmetry instead of against it. *Legion*'s legendary system, whose launch rule Chapter 18 takes apart, was repaired almost entirely by increases,[10] the direction that costs nothing on the day. A guarantee can also be sized so that it never becomes the reference. In a later season built from three older raids, players earned a currency over a number of weeks, capped at three, each unit buying one chosen item from those raids at Normal quality, below the Heroic and Mythic versions.[11] The developers called it an experiment, and the cap was there to protect against serious bad luck on a particular item without undercutting the main way of earning gear, playing meaningful content with a group over time.[12] Three

purchases, announced as a trial, are a floor under bad luck that nobody can mistake for a wage.

Holdings need not be useful to be missed. Pruning filler talents read to players as content lost for little gain, a case Chapter 2 examines.[13]

20.3 The Armor That Had to Be Earned Twice

Behind all of this sits a decision a game of equipment cannot avoid. Of the three routes for ***retirement*** in the report Chapter 18 reads from March 2002 (finite charges, a finite lifespan, or new items that are simply better), a game of gear progression lives mostly on the third.[14] Every new tier is then a managed devaluation of what players hold, carried out in the one form they accept: nothing is taken, and something better is offered.

Azerite armor broke that form. It was built to repair the system before it. In 2019 the game director, Ion Hazzikostas, described *Legion*'s artifact weapons as fantastic choice all at the front: two months of choices, then a straight line, and every player ended with the same fully unlocked weapon. Players had not complained; the designers saw it. Azerite armor was meant to supply interesting choices continually at the level cap, through armor whose traits the player unlocks.[15] The central complaint was that each new tier arrived with its traits locked, so a player who upgraded had to earn again what the old piece already gave, a step backward, "forcing you to re-earn the things you once had."[16] The better item was offered, and something was taken with it.

An interview six months earlier (October 10, 2018, against April 3, 2019) shows two more gaps. Players were at first flooded with Azerite from world quests at the level cap, then cut to a single weekly chance from a raid boss or a Mythic chest. And the choice was computed elsewhere: players modelled the traits until picking one became an externally solved math problem, in the sense of Chapter 5, and a piece whose trait the simulations rated poorly felt like no reward at all. The team had made more than 300 tuning passes on traits by hotfix, against the single structural change that followed.[17]

The same interview has a reference point set before release. Players had learned from the beta to expect three items from the weekly cache, the live game gave a different count, and the gap between reality and expectation was jarring even where the design was accepted.[18] This is ***expectation formation***: the reference moves up to whatever was received reliably, quickly, and also to whatever was shown in advance.

The repair took two patches. Tides of Vengeance added a second ring of traits and made traits easier to get. Rise of Azshara unlocked every trait by default and moved the earning into a separate item, the Heart of Azeroth. What players had been made to re-earn was handed over. The diagnosis was about process: the team waited too long to have the system fully working in the beta, and missed how players would move through the endgame, for there is "no substitute for throwing hundreds of thousands of players at an endgame system."[19] The repair ran in the cheap direction, after a loss with every tier had been built into the schedule.

20.4 The Cards Players Own

Magic retires cards by calendar, and Chapter 25 tells how Standard began amid the largest outcry the game had seen and how a reprint set angered the owners of the cards it reprinted. No card left anyone's box; a use and a scarcity did. A third version came with a new, scarcer rarity: many players hated a rarity that was harder to collect even though it made the others easier.[20] Rotation settles when cards leave, not what the next set must be worth, and Chapter 19 follows that promise into power creep. The collection is the reference: each set has to clear what players own without burying it.

The digital game designs around the loss directly. Dusting lets players destroy cards they own to make the ones they want, and the team behind *MTG Arena* declined it, believing that players keep the cards they open and earn, and that dusting turns the conversation from what to build into what to destroy, the more negative one. Players get wildcards instead, tokens that come with packs and can be redeemed for a card of a given rarity. The price is paid by players who want every card or build many decks.[21]

Loss works the same way inside one game. *Odyssey* was built to subvert card advantage, the rule that holding more cards than the opponent is good, and Patrol Hound let its owner discard from hand for a small combat ability. Most players did not like discarding their hands, and the lesson drawn was the difference between interesting and fun.[22] Players weigh the cards they give up more heavily than the effect they buy. Expectations carry from set to set as well: *Mercadian Masques* introduced three new groups of cards defined by creature type, with no keyword naming what they did, and players complained that the set had no new mechanics.[23] Earlier sets had taught them what a new mechanic looks like.

The asymmetry under every case here is what Kahneman and Tversky called reference dependence and loss aversion: outcomes are valued as gains

or losses from a reference point, and losses weigh more than gains of equal size.[24] Every case in this chapter fits it. As a framework it returns no number.

Design equation (a conceptual framework, not a law)

$$\text{felt change} = \text{new level} - \text{the level players are used to}$$
$$\text{intended change} = \text{new level} - \text{the level the design called for}$$

A reduction is in trouble wherever the first line is negative and the second is zero. The level players are used to includes one they were shown in a test build and never received.

20.5 When the Outcry Should Be Waited Out

The diary tells rest experience in another order. In May 2004 Pardo implemented a rest system that, instead of punishing players who played non-stop, rewarded casual players for taking sensible breaks. It was this system, described as a reward, that sent the forums berserk, with complaints that Blizzard was making a Care Bear game and forcing customers to limit their time. The hysteria passed in about a week, the bonus stayed after a couple of its values were tweaked, and by the following month it was no longer a controversy.[25] In the talk the protest came first and the relabeling ended it. In the diary the outcry greeted the bonus, players saw through it to the limit behind it, as Chapter 16 notes, and time and a small adjustment ended the argument.

A protest measures the reference point, not whether the reference deserves to win. The designers held their ground, the reference moved to the new system within weeks, and the game kept a rule the team believed in. Adaptation runs both ways, so some losses are worth imposing and waiting out.

The memory also defends more than generosity. The original game capped the number of debuffs on one enemy, a technical restriction of the earliest versions and no conscious design choice. When the studio rebuilt the old game as *WoW Classic* in 2019 it kept the cap for authenticity, and removed it only later, on one type of realm, calling the limits contentious.[26] What players are used to includes the walls.

The Designer's Test

1. What level of reward are players used to right now, and who set it: a designer, a bug, a test build, or a launch you wanted to go well?
2. If this number has to come down later, what will the reduction be measured against, and can the same schedule be shown from a lower reference?
3. Which rewards here are guaranteed, and how long before they are expected?
4. How does value leave this system, and in what form will holders accept the exit?
5. If players protest a change, what would tell you whether to wait a week or to move?

Notes

1. Frank Cifaldi. Report of Rob Pardo's talk Making a Standard: Blizzard Design Philosophies, Game Developers Conference. Game Developer, 2010, section "Make It A Bonus".
2. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "October 2003: Free Pizza and Other Hardships".
3. Cifaldi, report of Pardo's "Making a Standard", section "Make It A Bonus".
4. Staats, *The WoW Diary*, section "August 2004: Public Beta 4.0".
5. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V1 at 1:15.
6. Ornyx. Legendary Drop Rates. World of Warcraft Forums, `us.battle.net` (read from Wayback copy), 2017.
7. Blizzard Entertainment. Blizzard Insider 42: Meet the Pandaren Monk (interview with Greg Street). `worldofwarcraft.blizzard.com`, n.d., answer on death knight lessons.
8. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What didn't work out as planned?".
9. Street, "Cataclysm Post Mortem", section "What didn't work out as planned?".
10. Ornyx.
11. Chapter 14 lays out these difficulty tiers of one raid.
12. Blizzard Entertainment. [UPDATED AUGUST 2] Shadowlands Season 4 Content Update Notes. `news.blizzard.com`, 2022, section "Fated Raid Rewards".
13. Street, "Cataclysm Post Mortem", section "What didn't work out as planned?".
14. Staats, *The WoW Diary*, section "March 2002: Competitive Collaboration".
15. Steven Messner. Blizzard: There are 'things to learn' from the mistakes of Battle for Azeroth. PC Gamer, 2019, answer on the next expansion, paragraph 3.
16. Messner, answer on response time, paragraph 2.

17. Heather Newman. The State Of Warcraft: Ion Hazzikostas' Post-Mortem On The Battle For Azeroth Launch. Forbes, 2018, answers on priorities, paragraph 2, and on Azerite abilities, paragraph 1.
18. Newman, question "And then there was the confusion".
19. Messner, answers on response time and on the next expansion.
20. Mark Rosewater. Twenty Things That Were Going To Kill Magic. *Making Magic,* `magic.wizards.com`, 2013, item 14.
21. Wizards of the Coast. On the MTG Arena Economy in 2022. *MTG Arena,* `magic.wizards.com`, 2022, section "Player Collections".
22. Mark Rosewater. Twenty Years, Twenty Lessons, Part 1. Making Magic, `magic.wizards.com`, 2016, section "Lesson 5".
23. Mark Rosewater. Twenty Years, Twenty Lessons, Part 3. Making Magic, `magic.wizards.com`, 2016, section "Lesson 14".
24. Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, 1979.
25. Staats, *The WoW Diary*, sections "May 2004: The Care Bear Game" and "June 2004: Public Beta 2.0".
26. Blizzard Entertainment. Rules of Engagement: Classic Hardcore is Coming to World of Warcraft. `worldofwarcraft.blizzard.com`, 2023, section "To Buff or Debuff".

Part VI

Complexity

Chapter 21

Complexity Has a Budget

They made the cards easier to read, and the plan was carried out.

It started with a count. The people who make *Magic: The Gathering* found that fewer newcomers were learning to play, and they traced the cause to complexity creep. Every year's sets brought new cards, new mechanics and new concepts. Each gave an established player, who met them one at a time, a reason to stay. A newcomer met the whole pile at once, and it grew every year.[1]

The block then in production, *Time Spiral*, was too far along to fix, and its set *Future Sight* shows how many named mechanics one release had come to carry. So the repair went into the next block, *Lorwyn*, whose cards were to be straightforward, with clear rules text.[2]

The repair shipped. Then came the employee Prerelease for *Morningtide*.[3] The casual players in the building, the nearest thing the company had to the audience the simpler cards were written for, quit after one or two rounds.[4]

No card had beaten them. *Lorwyn* and *Morningtide* were full of cards whose effects depended on the other cards in play. Each was plain in the hand. Laid out side by side, a few of them multiplied what a player had to weigh before doing anything at all.[5]

The designers had spent a block lowering the cost of reading, and the players quit over the cost of looking up.

21.1 Three Costs on Three Schedules

The obvious repair is to make each piece simpler, and *Magic* had real failures of reading to justify it. Suspend, whose misreadings Chapter 1 followed, asked a player to hold more than one cost, an extra zone, counters and several turns of waiting. Testing showed that the term *the stack* confused beginners, so card text stopped using it, and the core sets printed reminder text beside even the basic keywords.[6] A card is also the unit a colleague

reviews: one card, one stranger, a clear pass mark. *Lorwyn* passed that test, and the load moved to other parts of the game.[7]

Rosewater's repair was to stop treating complexity as one quantity. ***Comprehension complexity*** is the cost of understanding what a card does when it is read. ***Board complexity*** is the cost of following how the cards in play affect one another. ***Strategic complexity*** is everything about using cards well: when to play one, how to spend resources, what to pick in a draft.[8]

A card that grows stronger for each other card of some kind in play costs almost nothing to read. Its cost repeats: every time the board changes, a player deciding whether to attack recomputes it, along with every card making the same kind of promise. Such cards become hard to track as the board fills.[9] Comprehension load grows with the number of cards, since each is read alone. Board load grows with the number of relationships among them, and relationships outrun cards quickly. No single card carried *Lorwyn*'s problem, so a per-card test could not catch it.

Under comprehension load players are confused, and the confusion need not come from hard text. Scornful Egotist was a creature costing eight mana in a set where high costs mattered to other cards. The wording was clear, and many players could not see why anyone would want it.[10]

Under strategic load, beginners do nothing unusual. The first two kinds of complexity are a problem for beginners and the third is not,[11] because strategic complexity has to be seen before it can be felt, and seeing it takes knowledge a beginner does not have yet.[12] A new player who casts a spell at the wrong moment is having a perfectly good game and has no idea it could be better. The depth costs nothing until the player can see it, and then it arrives as something to learn.

Attention is limited, the bounded capacity Chapter 8 traced to Herbert Simon,[13] and whatever is spent on tracking is not available for choosing. The three complexities are costs on different schedules. Comprehension is paid on first reading and falls with familiarity. Board complexity recurs at every decision and rises as pieces are added. Strategic complexity is deferred: nobody is billed until they can see it, and those who can count the payment as the game.

Design equation (a conceptual framework, not a law)

load on a player = comprehension (paid once, on reading) + board (paid at every action) + strategic (paid only by those who can see it)

The recurring term is the dangerous one. Comprehension load confused players. *Lorwyn*'s board load made them leave.

21.2 Rarity Decides Who Pays

Depth does not require visible complexity.

The tool that follows is a ***complexity budget***: a fixed allowance of the loads a beginner feels, spent on purpose and counted in total. Once R&D treated complexity at common as a limited resource, the goal became to move complexity to a different place in the game, not to take it out.[14]

The policy is New World Order, and its mechanism is rarity. Common is the most frequent rarity, and a beginner buys fewer packs than an invested player, so most of a beginner's cards are commons.[15] New World Order caps comprehension and board complexity at common, lets the higher rarities carry more, and leaves strategic complexity uncapped everywhere. The accounting has two levels. Each common is judged alone, and then all the commons are judged together, because a set's themes need some complexity at common and that portion has to be handed out deliberately.[16] The second level is the one *Lorwyn* lacked.

Rarity has a second job: deciding how often a player meets a card, it decides who meets the difficult ones. The result was better play, not merely easier play: R&D found the sets more fun, small effects that had to be tracked dropped out of the commons, and skill moved from keeping records to choosing well.[17] Bookkeeping had been mistaken for depth.

A cap says where complexity may not go. It does not say how to make a simple common worth an expert's attention. The answer is the ***lenticular*** card, named for printed images that change with the viewing angle: a card that reads as simple to a beginner and as deep to an expert.[18] Black Cat is a 1/1 creature that makes the opponent discard a card when it dies. A beginner sees a small creature with a bonus attached. An experienced player sees a trigger to plan a turn around.[19] Fact or Fiction is strong only in hands that can judge how cards are split and what the opponent is thinking.[20] Rescue from the Underworld returns a dead creature, and the way it works also takes one of the caster's creatures away and brings it back, which an expert exploits and a beginner never notices. The rule has two halves: cards that do not get in a beginner's way, and that give an advanced player options.[21]

A new kind of object draws on the same budget. Early in *Innistrad*'s design the team explored two versions of the double-faced card, one that transforms from front to back during play and one whose face the player

chooses when casting. The set shipped with the transforming kind only, tied to its theme of dark transformation, and the modal kind waited years for a home.[22] A brand-new physical object is itself an expense, and a player can learn one way of handling it at a time.

The budget is never settled. Reviewing the 2017 design year, Rosewater judged the sets too complex: *Kaladesh*'s limit of three mechanics had not been enough, and *Amonkhet*, with six keywords and many themes, had too much going on. The cut, he answered players who feared a dumber game, aims at information that rarely matters, not at depth.[23] A cap at common does not stop the total from creeping, because sets are built of mechanics as well as cards. The advice on combining them is the *Lorwyn* lesson restated for a set file: the best second mechanic plays in a different space from the first.[24]

21.3 Four Buttons and an Auctioneer

World of Warcraft keeps the same accounts with different line items. Chapter 2 reported the count that sets the scale: about four abilities in a regular rotation is the sweet spot, against original classes that leaned on one main ability and carried many unused ones.[25] What belongs here is why the count only rises. New emergency buttons for healers crowd out a core spell, most often Flash Heal.[26] Players expect new abilities at each new level cap, and designers feel pressure to repair a neglected ability when cutting it would be wiser.[27] The cost falls where it fell in *Magic*: a player returning to a Fury warrior after time away has to learn the rotation again, and depth that interests a veteran can be incomprehensible to someone new or returning.[28]

The cut came in a later post. For years more abilities had been added than removed, and some players felt they needed dozens of keybinds. The aim was understanding without less depth, and the methods were to restrict some abilities to one specialization, remove others, fold redundant passives into the abilities they modified, and merge or remove cooldowns typically used together.[29] Dozens of keybinds were a large number that described very little. A count of options is not a count of decisions, the argument Chapter 2 took from Rosewater's creature that gains an ability and no interest.[30]

The *Lorwyn* lesson, that the load which hurts is aggregate, shows in crowd control.[31] The same post says there was simply too much of it, above all between players, and the fix worked on the total: silences were detached from interrupts, many control spells were removed, and the categories sharing diminishing returns were merged into fewer.[32] Talents show it in choices.

Many small incremental talents with a mathematically right answer are not a good model.[33] Such a talent has to be read and is settled by a spreadsheet: load with nothing to decide.

A new class shows a team budgeting in advance. Looking back on the death knight, Street said its resource model had ended up feeling very complicated and that the monk would get a simpler one. For the monk's memorable early ability, the role Death Grip had played, the team chose Roll, a burst of movement between the warrior's Charge and the mage's Blink.[34] A new class has only so much novelty a player can absorb in the first hours. The death knight spent it on a resource system to be studied, the monk on one action understood by pressing it once.

Depth can also be kept whole and moved out of sight. Rob Pardo's example was the auction house, a whole economy reached by walking up to a character in the world and speaking to it, with no button on the screen. Only key information belongs on the heads-up display, and putting everything on the interface is a common pitfall.[35] The auction house costs a new player nothing until the day that player goes looking for it, which is the property of strategic complexity. The intent went back to the start: Allen Adham liked to use chess to show how a simple game could be played at a higher level of complexity. "Everything was measured in terms of intuition and ease of play."[36]

Rarity sorts players by packs opened, a specialization by what they chose to play, an auctioneer by who walked over to ask. The full system exists for everyone, and each player is charged for the part reached. This is segmentation: one product arranged so that different buyers meet different parts of it, as the difficulty tiers of one raid did in Chapter 14.

Time inside one game is a fourth sorter, and it needs no collection. A game of *Magic* starts with almost nothing on the table and reveals itself a piece at a time, so the opening turns work as a tutorial. Within a turn, once attackers are declared, a defender knows which kinds of card can no longer arrive. The hosts of the design channel Distraction Makers call this chunking, making information digestible.[37] A board that fills slowly sends the recurring bill late.

Second-Order Effect: The bill for simplifying

Both games paid for their first simplification. In *Magic* clearer text led to cards that leaned on one another, and the cards written to be easy were the ones casual players walked away from.[38] In *World*

> *of Warcraft*, *Cataclysm* pruned filler talents, and Chapter 2 tells what followed: one acclaimed part aside, the overhaul felt like a bad bargain, content lost for little gain, and talents were redesigned again for the next expansion.[39] Removing a false choice leaves a gap where players used to have something, and they notice the gap before the clarity.

21.4 Where Hidden Load Comes Back

A simple surface can mislead the person it was simplified for. Venerable Monk, the life-gaining creature of Chapter 1, is a success because a beginner is happy to play it. Aven Cloudchaser is the counterexample: a 2/2 flier that destroys an enchantment when it arrives. A beginner reads the bonus as the point, holds the creature waiting for a target, and grows frustrated at not having the flier in play. A lenticular card has to produce a good game at both ends of the range of players.[40] The default play of someone who cannot see the depth has to be a decent one. The exemption also has a clock: board complexity is invisible to beginners only at first, and they soon start looking at how cards interact.[41]

Cutting is a loss to someone. The pruning post acknowledges that players feel strongly about their abilities and asks those whose favorite was removed to understand why.[42] The budget is the designer's ledger. The player's ledger holds a button that sat on the bar for years.

Players can also move load off the designer's books. *World of Warcraft* add-ons read the state of a fight and compute things for their users, and Chapter 15 follows how they came to solve fights. The current team is restricting the data add-ons may use in combat and simplifying classes to compensate.[43] Complexity handed to a tool returns as an assumption about what every player runs, and a new player without the tools faces a game tuned for those who have them.

A paper game has no add-ons, and its load moves onto the table. Distraction Makers call the components a game needs beyond its own systems cruft, and their plainest case is the life total, kept on a die or a phone. A component earns its place only if removing it removes a function. The danger is state that a card asks players to track and gives them no way to see, such as an ability allowed to trigger only a set number of times each turn: players invent a tracker, and triggers are missed.[44]

The Designer's Test

1. Of the complexity in this piece, how much is paid when reading it, how much every time the player acts, and how much only by experts?
2. Who meets this piece first, and how often will they meet it?
3. If this text gets simpler, where does the load go?
4. What will a beginner do with this piece by default, and is that a good game?
5. What could be removed that nobody would miss, and what would be missed although nobody uses it?

Notes

1. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "Problem Solving".
2. Rosewater, "New World Order", section "A Complex Problem".
3. A Prerelease is an event at which players build decks from a new set before it goes on sale. This one was held inside the company.
4. Rosewater, "New World Order", section "Board Complexity".
5. Rosewater, "New World Order", section "Board Complexity".
6. Mark Rosewater. Lenticular Design. Making Magic, `magic.wizards.com`, 2014, section "Comprehension Complexity".
7. Rosewater, "New World Order", section "A Complex Problem".
8. Rosewater, "New World Order", section "A Complex Problem".
9. Rosewater, "Lenticular Design", section "Board Complexity".
10. Rosewater, "New World Order", section "Comprehension Complexity".
11. Rosewater, "New World Order", section "Common Knowledge".
12. Rosewater, "Lenticular Design", section "It Depends on How You Look at It".
13. Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.
14. Rosewater, "New World Order", section "Common Knowledge".
15. Rosewater, "Lenticular Design", section "It Depends on How You Look at It".
16. Rosewater, "New World Order", section "Common Knowledge".
17. Rosewater, "New World Order", section "A Whole New World".
18. Rosewater, "Lenticular Design", section "It Depends on How You Look at It".
19. Rosewater, "Lenticular Design", section "It Depends on How You Look at It".
20. Rosewater, "Lenticular Design", section "Strategic Complexity".
21. Rosewater, "Lenticular Design", section "Rule 6".
22. Mark Rosewater. Study Session. Making Magic, `magic.wizards.com`, 2026, section "A Modal Double-Faced Card (MDFC) Set".

23. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section "We were too complex".
24. Mark Rosewater. Nuts & Bolts 18: Layering Your Mechanics. Making Magic, `magic.wizards.com`, 2026, section "The Second Mechanic".
25. Greg Street. Dev Watercooler: Number of Abilities. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What's the Magic Number?".
26. Street, "Number of Abilities", section "What's the Magic Number?".
27. Street, "Number of Abilities", section "The Human Factor".
28. Street, "Number of Abilities", section "Too Many to Handle?".
29. Blizzard Entertainment. Dev Watercooler: Pruning the Garden of War. `worldofwarcraft.blizzard.com`, n.d., section "Ability Pruning".
30. Mark Rosewater. Decisions, Decisions, Part I. Making Magic, `magic.wizards.com`, 2009, section "Level I, The Card Level".
31. Crowd control is any effect that takes away a character's ability to act or move. Diminishing returns here means that repeated control effects of one category on the same target last a shorter time on each use.
32. Blizzard, "Pruning the Garden of War", section "Crowd Control and Diminishing Returns".
33. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What didn't work out as planned?".
34. Blizzard Entertainment. Blizzard Insider 42: Meet the Pandaren Monk (interview with Greg Street). `worldofwarcraft.blizzard.com`, n.d., answers on death knight lessons and on Roll.
35. Mark Wallace. Report of Rob Pardo's keynote on World of Warcraft at the Austin Game Conference. Game Developer, 2006, paragraphs 3 and 4.
36. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "Antecedents: Nomad and Warcraft III".
37. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 4 at 9:37; episode 19 at 5:19 and 6:28.
38. Rosewater, "New World Order", section "Board Complexity".
39. Street, "Cataclysm Post Mortem", section "What didn't work out as planned?".
40. Rosewater, "Lenticular Design", section "Rule 5".
41. Rosewater, "Lenticular Design", section "Rule 1".
42. Blizzard, "Pruning the Garden of War", section "Ability Pruning".
43. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, sections "Addons & the Trickle Down Effect" and "Opening Up Encounter Design".
44. Distraction Makers, episode 9 at 0:07 and 4:46; episode 19 at 15:29, 16:38 and 21:23.

Chapter 22

Design the System, Not the Feature

Karazhan is a wizard's tower, and it was built three times before anyone fought in it.

John Staats built it first, over six months in a level editor called Radiant. Then the team moved dungeon building to 3D Studio Max, and the tower started again from nothing.[1]

A year later a second level designer, Jose Aello, took it up. Following Chris Metzen's direction that the tower be very tall and slender, he spent a few months on it and presented a fully textured version to the game designers, who had not seen it before. No game designer was in charge of dungeons, because dungeons were not yet playable. The designers found every room too narrow for combat and half the floor area given to stairs, and the whole tower was scrapped for lack of playable combat space.[2]

A year after that, a morning-long meeting produced a tower ringed by smaller towers floating free of it, and the programmers ended the idea because the pathing data could not support complicated moving structures. Months later Karazhan returned to Staats with a clear instruction: a raid dungeon as big as upper and lower Blackrock Spire combined. When the request was made, "we hadn't played a raid yet." His plan was so large that a colleague built the interiors while he built the shell, with a second raid on top, flooded levels underneath and cellars through the town below. Ordered to reduce it, the builders cut the flooded levels, the upper raid and the cellars, and it was still too big.[3]

By then the team had discovered that marathon dungeons were not fun, and for the first time anyone called a dungeon too big. The layout was reduced again before the tower opened in the game's first expansion, in 2007.[4]

None of the three towers was badly made. Each answered the question as posed, and each time the question came before anyone knew what a tower in this game was for.

22.1 Every Piece Waits on Another

A game the size of *World of Warcraft* is built in parallel, by specialists, and the systems they depend on arrive in an order. On Staats's first day Eric Dodds showed him one character on the terrain with an empty loot table: movement had come first, and combat, items and the interface would follow.[56]

"The paradox was that artists, producers, and programmers tried to support gameplay before it actually existed." The team worked long hours on educated guesswork, with no answer to how crowded zones would be, how large a party should be or how long a dungeon run should take.[7]

Call that a ***dependency***: a piece cannot be judged until some other piece is working. Waiting was not an option, because a dungeon took between two and nine months to build, and texture artists might take as long again to paint it.[8] Dungeons could not be reviewed inside the game, so neither art nor design leads could easily follow them, and some were rejected after months of work.[9] Done meant finished, not tested, because nothing existed to test against.

The missing piece was the monster. In October 2002 the pathing code might have to be rewritten from scratch again, and the delay meant the designers still could not tell level designers how to build dungeons.[10] The next month, in some areas, creatures would not move at all.[11] Stable indoor pathing arrived in June 2003 after a year of on-again, off-again work. Only then could designers place monsters and script fights, and testing group combat indoors might redefine how dungeons should be built.[12]

Wailing Caverns carries the marks of that gap. After its second iteration the designers wanted it a lot bigger. Nobody could yet test fighting there, so Staats attached two enormous wings and a maze to see whether mazes held something interesting. They did not, and the trail of mushrooms he laid to mark the shortest way vanished into the clutter.[13] The section of the diary that tells this carries a line from Jeff Kaplan at its head: "There's no such thing as a dungeon that's too big. When in doubt, make it bigger." Without answers, the team built things to see what felt right.[14]

Play corrected the scale. Big dungeons were the inherited pattern: in the older game the team knew, a dungeon was a shared space that had to

hold everyone on the server, and the team feared players would run through content too fast. Nobody foresaw how much private copies of a dungeon would shrink the space needed, and a run of about an hour did not exist in the genre. Scarlet Monastery, which Aaron Keller divided into wings, came nearest, and only after launch did the team learn that players loved short runs. Wings became the direction for later expansions.[15] When friends-and-family testers played the Deadmines late in 2003, the team finally knew how long a good crawl should be, and many dungeons were then enlarged to match it.[16] The measurement needed a finished dungeon, working monsters and people from outside the team.

22.2 An Owner, a Tool, a Multiplier and a Test

A good mechanic in isolation does not guarantee a good game.

The second Karazhan was a good tower and a useless dungeon, and the difference lay outside the tower, in a combat system that needed floor. So the thing to design first is the set of roles and their connections: what a dungeon is for, how long a group stays, what a quest does for a zone. Pieces fill roles and are judged by the role's owner before the art is finished. The second tower had an author and a lore brief, and no owner, because its role did not yet exist.

The missing structure had four parts. The first is an owner. Before *World of Warcraft* the same team spent eighteen months on Nomad, a project of programmers and artists whose approach to design was democratic, too democratic. With nobody in charge of design, ideas did not connect and compromises piled up until the project made no one happy.[17] The new project put Allen Adham in charge of design, with designers who heard every argument and owned the final decisions, and gave the team *EverQuest* as a common frame of reference. It broke more ground in three months than Nomad had in eighteen.[18][19] An owner and a shared reference are structure too.

The second is a tool. Until the spell editor existed, all combat was fake: a sword blow dealt no particular kind of damage and monsters had no armor, because any change to a number meant asking a programmer to hardcode it. Designers specified a database, programmers built it, other programmers connected it to the game, and only then was a sword a sword. Group combat took about nine months more of spell work.[20] The editor became usable in July 2002, and content creation lurched into overdrive.[21] Designers built 99 percent of the game's abilities themselves, and only a couple needed a

programmer.[22] The most valuable system in a project can be one no player sees: the one that lets the others be evaluated.

The third is a cost that multiplies. Nine player races were intended, with demons, goblins and naga joining the six approved.[23] A race is not one asset. Every race wears every piece of armor: each shield was resized, reshaped and indexed sixteen times, once for each race and gender.[24] Naga have a tail, ruling out leggings, boots and capes. Demons were meant to change shape. Goblins needed a world of machines that could not be reused. The three would easily have tripled the work of the other six combined, and they were dropped.[25] Late in 2002 artists were still failing to find a naga body that could wear armor in parts.[26] As a feature, a snake-bodied race is an easy yes. As a part, it is a claim on every item anyone will ever make.

The fourth is a test applied before building: will players want to do this every day, and can the feature be reused to branch into other kinds of play?[27] Staats's bounty pitch (Chapter 3) failed it, and so did unique artifacts out of reach for 99.99 percent of a server. The team built only reusable features, with two named exceptions, the fishing mechanics and the talent interface.[28] The question asked of a feature was what else it fed.

Second-Order Effect: More dungeons, fewer players in each

Chapter 2 tells how an endless supply of generated dungeons won the designers over in July 2001 and lost to thirty hand-built ones by October, partly because too many dungeons might divide the player base in a social game.[29] Judged by its role, each added dungeon was a gain for the group standing in it and a loss for every group trying to form somewhere else.

22.3 The Skeleton Before the Cards

A *Magic: The Gathering* set is a few hundred cards, each printable on its own, and it begins as a ***design skeleton***: a list of slots by color, rarity and card type that says what the set needs before any card exists.[30] In Rosewater's worked example, an imaginary set with 60 commons, about half the slots are creatures, and the five colors divide those 30 by their usual shares.[31] The skeleton is the roles, written down first.

It shows what has to be allocated, and the usual trouble is too much to fit.[32] Once slots go to artifacts and lands, white has no room for all its creatures, so one creature is cut and the need is met by a spell that

makes creature tokens.[33] The skeleton is filled from the commons up and the hardest slot first, because each choice narrows the next, and the set matters more than any card: a designer in love with one card is what most likely derails it.[34] At assembly, a card that does not advance the set goes however good it is, and can be saved for another.[35] Nothing is judged alone.[36] The second Karazhan was a card that met its own brief and did not advance the set.

The same error exists one level up. In an older process the first set of a block was designed while the plan for the whole block was still being worked out.[37] Small decisions in the first set became constraints on the block, later developers had fewer options, and there was little time to reflect. The repair was a planning stage before design, whose job is learning what the questions are, not answering them.[3839]

A skeleton also makes room for parts no player would list among a set's attractions: a spread of costs so there is something to play each turn, evasion so games end, answers at every rarity so Limited games are not decided by one powerful rare.[40] This is ***structural support***. A mechanic can be designed purely to provide it: one set had a token theme and a counter theme with nothing connecting them, and a mechanic was built to tie the two together.[41]

Support can sit above the cards altogether. The design channel Distraction Makers separates a game's systems level, the rules outside the cards, from its mechanics level, what the cards do. The turn gives a player one main phase before combat and another after, so any card can be played in at least two ways, and the way a newcomer finds natural, everything before the attack, works without being the best line most of the time.[42] One structural rule does on every card the work a feature would have to be printed to do.

The opposite of a connecting mechanic is a parasitic one, which works mostly with its own kind: early poison in Chapter 10,[43] energy in Chapter 23.[44] It can be fun and still fail as a part, because it feeds so little outside itself.

22.4 The Order of Construction Is a Design Decision

A piece cannot be validated before the systems it needs, so whatever is built first is built on a guess, and whatever is built last decides when the guesses get checked. In *World of Warcraft* movement came first and pathing and

the spell editor late, which helps explain why most design decisions came at the tail end of the project, a pattern common for games built on their own engine.[45] The team even had to choose between dependencies. In October 2001 the dungeon builders asked for a tool to place props indoors and were refused, because the same programmer was writing the spell editor and combat testing was waiting on it.[46] The exposure is integration risk: parts that each pass their own inspection fail when they first meet, at a cost that rises with how finished the parts are. The second tower was fully textured (Figure 22.1).

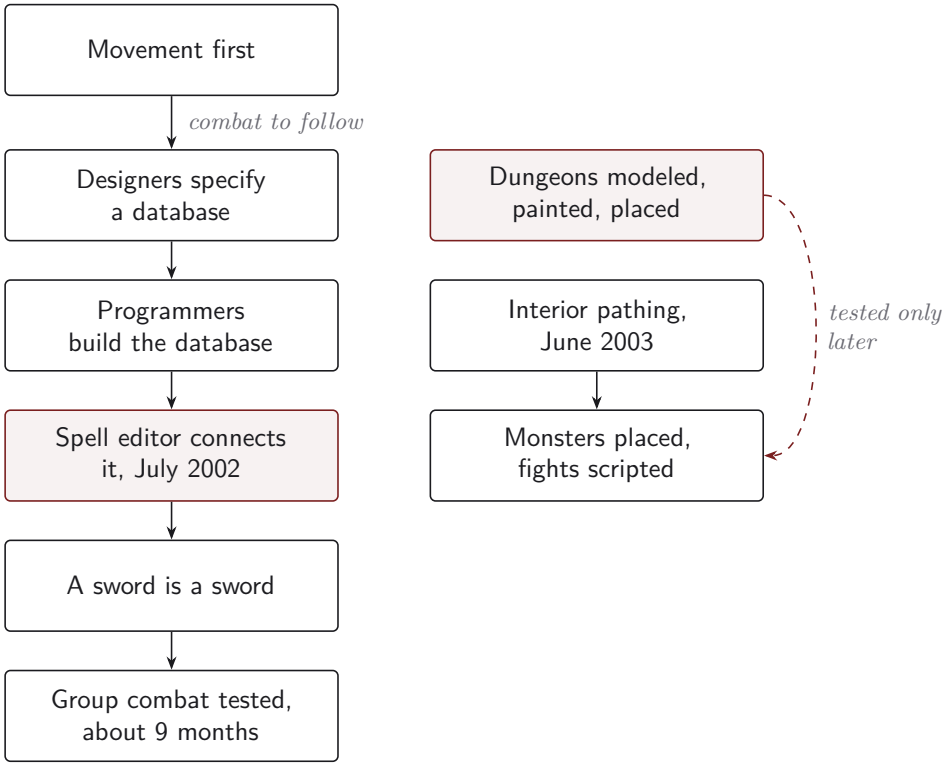


Figure 22.1. What had to exist before a dungeon could be judged, from Staats's diary. The dashed arrow marks dungeons finished before anything could test them.

The gap between local and global optimization, which Chapter 1 applied to players, applies to makers too. The tallest tower, the most beloved card and the most polished first set each answer a local question best and can

leave the whole worse. Both studios gave someone the global question as a job, with power to overrule the local best.

Structural support suffers at the review table. Nobody prints confirmation buttons for trades on the back of a box, and details of that kind kept the game fun.[47] A review that asks only whether each piece is exciting cuts the supports first.

22.5 When the Play Overrules the Plan

The largest success in this chapter was not in the architecture. Quests were meant as signposts, and players made them the way to level (Chapter 1). The team followed them. The producers had expected to write quests in their spare time and ended up hiring more than half a dozen full-time quest designers, a department the plan had no role for.[48] A team that held quests to their assigned role would have shipped the grind.

Staats argues against this chapter's title. Designers should build from solid moment-to-moment play and discover where it leads, and it is the inexperienced who cling to grand schemes of interrelated systems before finding out whether they are enjoyable.[49] Dodds held that detailed documents were a poor way to design while gameplay was unsettled.[50] Naming the roles first is not specifying the pieces first, and Karazhan asks for a named owner and an early look, much less than a master plan. The skeleton bends the same way: it does not exist to stop a designer finding other things, and when a good card turns up that it did not anticipate, the skeleton changes.[51]

A skeleton also presupposes a frame somebody already knows. The *World of Warcraft* team borrowed its frame from an existing game, a borrowing Staats calls crucial.[52] Where the frame had no answer, as with the length of a private dungeon the older game never had, the pieces built in the meantime were the ones rebuilt. A team making something with no predecessor cannot write the skeleton first. It can keep the pieces cheap until the structure shows itself, and a fully textured second Karazhan was the opposite of cheap.

The Designer's Test

1. What does this piece need to exist before anyone can tell whether it works?
2. Who owns the role this piece fills, and have they seen it yet?

3. Is progress here being counted in pieces finished or in pieces validated?
4. What else does this feature feed, and what does it claim from every piece made after it?
5. Which part of the plan is the play already contradicting?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "Dungeons: The Last Hurdle".
2. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
3. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
4. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
5. John Staats. Making WoW: Excerpts from The WoW Diary and AMA with Author John Staats. Wowhead, 2019, paragraph beginning "Eric interpreted my exclamation".
6. Staats, "Making WoW: Excerpts", paragraph beginning "Eric gave me a brief history".
7. Staats, *The WoW Diary*, section "Gameplay".
8. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
9. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
10. Staats, *The WoW Diary*, section "October 2002: Still Unanswered Questions".
11. Staats, *The WoW Diary*, section "November 2002: Internal Alpha 2.0".
12. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
13. Staats, *The WoW Diary*, section "March 2003: Internal Alpha 3.0".
14. Staats, *The WoW Diary*, section "March 2003: Internal Alpha 3.0".
15. Staats, *The WoW Diary*, section "March 2003: Internal Alpha 3.0".
16. Staats, *The WoW Diary*, section "December 2003: Stepping on Toes".
17. Staats, *The WoW Diary*, section "Antecedents: Nomad and Warcraft III".
18. Staats, *The WoW Diary*, section "Antecedents: Nomad and Warcraft III".
19. John Staats. Making WoW: Learning from the Mistakes of Nomad. Wowhead, 2019, last paragraph.
20. Staats, *The WoW Diary*, section "Gameplay".
21. Staats, *The WoW Diary*, section "July 2002: A Modicum of Luster, A Pivotal Juncture".
22. Staats, *The WoW Diary*, section "Wowedit".
23. Staats, *The WoW Diary*, section "Character Design".
24. Staats, *The WoW Diary*, section "Animation".
25. Staats, *The WoW Diary*, section "Character Design".
26. Staats, *The WoW Diary*, section "November 2002: Internal Alpha 2.0".
27. Staats, *The WoW Diary*, section "Why MMOs Are So Difficult to Create".
28. Staats, *The WoW Diary*, section "Announcement at the ECTS".
29. Staats, *The WoW Diary*, section "October 2001: Learning from the Good and Bad".
30. Rarity is how often a card appears in packs, common being the most frequent. Limited is the way of playing in which decks are built from freshly opened packs, so the commons are most of what a player sees.
31. Mark Rosewater. Nuts & Bolts 2: Design Skeleton. Making Magic, `magic.wizards.com`, 2010, sections "#1" and "#2".

32. Rosewater, “Design Skeleton”, section “Putting in all the keywords”.
33. Rosewater, “Design Skeleton”, section “Let’s layer this in”.
34. Mark Rosewater. Nuts & Bolts 3: Filling In the Design Skeleton. Making Magic, `magic.wizards.com`, 2011, section “Step #4”.
35. Mark Rosewater. Nuts & Bolts 7: The Three Stages of Design. Making Magic, `magic.wizards.com`, 2015, section “It Sheds Unneeded Elements”.
36. Mark Rosewater. Nuts & Bolts 9: Evaluation. Making Magic, `magic.wizards.com`, 2017, section “Question #2”.
37. A block was a group of consecutive sets sharing a setting and mechanics.
38. Mark Rosewater. Advanced Planning. Making Magic, `magic.wizards.com`, 2014, section “What Is Advanced Design?”.
39. Rosewater, “Advanced Planning”, section “Before:”.
40. Mark Rosewater. Nuts & Bolts 15: Structural Support. Making Magic, `magic.wizards.com`, 2023, sections “Mana Curves”, “Evasion” and “Threats & Answers”.
41. Mark Rosewater. Nuts & Bolts 18: Layering Your Mechanics. Making Magic, `magic.wizards.com`, 2026, section “Necessary Elements”.
42. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 4 at 3:42 and 4:14.
43. Mark Rosewater. Something Wicked This Way Comes, Part 3. Making Magic, `magic.wizards.com`, 2010, section “Problem #3”.
44. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section “Too parasitic”.
45. Staats, *The WoW Diary*, section “Announcement at the ECTS”.
46. Staats, *The WoW Diary*, section “October 2001: Learning from the Good and Bad”.
47. Staats, *The WoW Diary*, section “Gameplay”.
48. Staats, *The WoW Diary*, section “Quests”.
49. Staats, *The WoW Diary*, section “Gameplay”.
50. Staats, *The WoW Diary*, section “Announcement at the ECTS”.
51. Rosewater, “Filling In the Design Skeleton”, section “Step #5”.
52. Staats, “Mistakes of Nomad”, paragraph beginning “Eric Dodds had been promoted”.

Part VII

Iteration

Chapter 23

Most Good Mechanics Begin as Bad Mechanics

The first dungeons of *World of Warcraft* were built in Radiant, a level editor made for first-person shooters and chosen for its mathematically clean geometry. It did not wear well. The artists disliked its rigid grid, finished work was very hard to resize, and John Staats's Karazhan file took over five minutes to save in a program that kept crashing.[1]

None of that ended it. In July 2001 the programmers found that Radiant geometry could not be streamed, loaded piece by piece as a player approached. Every city and building made in it would have needed a loading screen, in a world where nothing else used one. The interiors were scrapped, and the team moved to 3D Studio Max, the program it already used for creatures and props.[2]

The new tool repaid the move at once. Staats modeled a whole dungeon in it in one week, where the old unit of work had been months.[3] Two months later he called the switch crucially correct.[4]

Look at what sank the editor. The grid, the resizing and the crashes were costs, and a team pays costs every day. The deciding item was of another kind. It was a property of the game Radiant was being used for: a world meant to be crossed without a loading screen could not be built from pieces that loaded whole. Nothing on the list says the editor was bad at what it was made for. It says the editor and this game could not live together, and it says exactly where they collided.

23.1 The Verdict and the List

Every early judgment is made in a setting that is missing or borrowed. Radiant was chosen before anyone had tested whether its geometry could stream, and the buildings made in it met the same trap. The game designers, new to 3D, needed to run through finished levels before they could say

how high ceilings had to be, how much floor a fight needed, or whether layouts should be linear, and a dungeon file took hundreds of hours to build. Staats calls it a chicken-and-egg paradox.[5] Karazhan, sized before anyone had played a raid, is the result Chapter 22 tells.

The usual answer is to prototype, play and keep what is fun. It is a good rule, because it replaces argument with a test. Blizzard's founders had instilled the philosophy of fixing whatever did not work and redoing it until it was as flawless as possible.[6] It also keeps a ledger with one column: an idea in, a verdict out.

The verdict on Radiant had two parts, and the rule reads only the first. The first is *wrong tool*. The second is the list, and its deciding item is a property of the host game. The grid and the slow saves belong to the editor. The loading screen belongs to a world built to be crossed without one. The sentence *Radiant does not work* is short for a longer one: Radiant does not work for a game whose world has to stream.

The long sentence is the one that can be reused. Its condition can be checked against any later tool in minutes, and it tells the team what the replacement has to do before anyone shops for one. The short sentence can only be filed, and the loss is silent, because nobody reopens a rejected idea to ask what the rejection depended on.

The rule fails a second way when the thing under test never reaches the players. *Diablo III*'s set monster, a pack of four enemies in which each hands its ability to the others as it dies, was built at Blizzard North in 2004 and judged not fun. Players there killed everything at once with area attacks, did not read effect icons and were not talking, so nobody noticed the transfer.[7] A test in which the subject goes unnoticed has measured the room. Its designer carried those three reasons to the Ahn'Qiraj raids, found a game in which each was reversed, and built the same pack as the Anubisath Sentinels, the case Chapter 12 tells.[8] An idea like set monsters "isn't good or bad on its own."[9]

A pass or a fail carries one bit, and it describes one pairing of idea and setting. *Wrong tool* told the team about one editor. *It cannot stream* told it what every tool after Radiant had to do. The practical difference between the two kinds of test is small: someone asks why, and someone writes the answer where it will be found.

A failed prototype is worth the reasons it failed. The reasons travel to the next attempt and the next system, even when the mechanic stays behind.

Call the list the ***conditions of failure***: the properties of the host system under which the idea does not work. They are the output of a prototype worth keeping, more than the verdict and often more than the prototype. A verdict closes a question. A condition leaves it open in a precise place, because it says what would have to be true for the answer to change. What separates a good mechanic from its bad earlier version is often a reason somebody could state.

23.2 Reasons That Traveled

Adham's month of diminishing returns in place of cooldowns, which Chapter 4 tells, failed with its reason attached: players under pressure needed hard stops and visual cues.[10] Because the reason was about players and not about that combat model, it applies to fights designed later by other people, and the introduction's Loatheb team reached the same one separately. The tests ran on the ability editor, which let designers build and balance encounters without a programmer.[11] The combat model was thrown out and the bench it ran on stayed.

The talent system ran the loop three times, and Chapter 2 follows its versions as menus. Read as prototypes, each was built against the stated failure of the one before. The first let a player pick a stat to raise, and every melee class picked the same stats. The second offered specializations in a weapon, a school or a foe, and failed because even an odd choice left the game playing exactly the same. The third, a tree for each class, was aimed at that sentence, at about a hundred times the work of the original, and made one player's experience different from another's in the same class.[12] The failures got more exact: first players chose alike, then the choice changed nothing they did.

The editor of the opening shows what a late reason costs. This is ***path dependence***: early choices make later ones cheaper or dearer, so the order of decisions shapes the outcome. Radiant was a reasonable first choice, and every month of interiors built in it raised the price of leaving. A rational team ignores sunk costs when choosing, and this team did. It could not ignore the write-off, which is real even when the decision that causes it is correct. Path dependence is the reason to want conditions of failure early, while acting on them is cheap.

Second-Order Effect: The price of being right

Everything built in the abandoned editor went with it, including Staats's six months on Karazhan.[13] The decision came just after the team hired its first professionally experienced 3D level designer, who moved to California to find his new job would use none of his Radiant expertise. The month's title in the diary is nine months down the tubes, and its explanation is that it sometimes took a long time to figure out that something would not work.[14] Every month it took added a month of interiors to the pile that would be thrown away.

One prototype ran the other way. Flying was not in the game. In August 2002 the producer Shane Dabiri used a cheat to swap his character's model for one seated on a wind-rider and turned off gravity. Each jump lifted him six feet and left him there, until he was running above the rooftops, wings flapping, unable to glide up or down. The team gathered at his desk anyway and began talking about air taxis between zones.[15] The engineering cost came after: flying taxis created severe frame rate problems, and flight paths were restricted to areas where the drop would not be noticed.[16] By April 2003 the team was looking forward to announcing flying taxis at E3, because they set the game apart.[17] A prototype can show why something works before anyone can build it, and the conditions it reveals (which routes, at what frame rate) shape the feature as conditions of failure do.

23.3 Energy, Chroma and the Rules for Failing

Magic: The Gathering has its own shelved idea in energy, and Mark Rosewater wrote down both of its rejections. In the original *Mirrodin* the mechanic tested well and produced interesting choices, and it made players track which artifact held how many counters, a great deal of mental strain for very little gameplay. He reworked it, and it was cut anyway: the set was overstuffed, and energy was less woven into the rest of it than the other mechanics were. In *Shards of Alara* it was pulled again, an imperfect fit that needed more room than the set could give. Rosewater kept looking for a home for it because he thought it was a good mechanic.[18]

Neither rejection said the mechanic was bad. Both said something about the set: no room, too little connection. *Kaladesh* reversed those conditions on purpose, and the first prototype's reason was still at work there. Of the three versions tested, the team passed over the one that kept counters on

cards, because it took more bookkeeping without better play. The version chosen was the one Rosewater had designed first, a fact he kept from the testers so it would not sway them.[19]

Energy was held between sets the way Chapter 4 holds a spell on cooldown: option value, an action worth more unspent because spending it gives up a better moment. A rejected idea with its conditions attached costs nothing to keep and waits for a host that can afford it. Good cards cut from a set can be saved for a future one,[20] and *World of Warcraft*'s team shelved player housing for the first release until someone figured out how to make it fun.[21] An option nobody remembers has no value, which is the argument for a written shelf.

Chroma separates the idea from its execution. It counted colored mana symbols, was received as lackluster, and Rosewater counts it a mistake. Devotion, made for a later set, counts the same symbols, repaired two things chroma lacked, a flavor and something consistent to build a deck around, and was a large success. Even the best idea can fail when it is not designed correctly.[22]

The studio has written its rules for failing down. Some cards failing in playtest does not mean their mechanic or theme should go: the designer finds the aspect that works and moves more cards toward it.[23] Early playtests use small decks of simple cards built around the new mechanic, and the testers are asked whether it was fun and, if not, why not. Fixable mechanics are adjusted and tested again, and the rest are dropped.[24] A first mechanic that stops delivering can be replaced, and moving backward is an ordinary part of design.[25] Some failures hold no lesson about fit at all: R&D simply misjudges, as with a cycle it expected to succeed in multiplayer games that did not.[26]

A reason is worth something only while the thing it describes can still change. The complexity problem of Chapter 21 was identified while *Time Spiral* was too far along to alter, so the repair went into *Lorwyn*.[27] A test whose result arrives after the design is fixed teaches the next project, and the current one ships as it is.

23.4 Not Every Bad Mechanic Is Waiting

Fitting the home is not fitting the world. Energy worked in the set built to hold it. A year later it was listed among the lessons as too parasitic, its cards wanting a deck full of their own kind.[28] The conditions from two earlier sets had been met in full, and meeting them produced a mechanic

that sat badly with everything outside its own set. Reasons drawn from past hosts say nothing about the costs a new host will discover.

Iteration is paid in calendar time. The talent design was declared finished in July 2003, and it was not.[29] The tree that fixed it was under way the following June, with a month spent on the first two of nine classes and three months left before the gold master.[30] The third attempt was the good one, and it nearly did not fit. A team with one fewer designer or one fewer month would have shipped the second version, reasons and all.

Some reasons close the question for good. The generated dungeons of Chapter 2 lost their own supporters within a week of a counter-demonstration, tired of recognizing the same rooms and missing a sense of place.[31] Those reasons were about what the game was for, and no later context reversed them. Mounted combat went the same way: nobody found a workable version, and enthusiasm fizzled out.[32] The title's claim concerns where good mechanics come from. It makes no promise about where bad ones go.

The Designer's Test

1. When this prototype failed, what exactly failed: the idea, or something about the place where it was tried?
2. Did the testers ever meet the mechanic, or did the test measure the game around it?
3. What would have to be true of a game for this idea to work?
4. What is on the shelf, and is the reason it is there written where the next team will find it?
5. Will the result of this test arrive while the design can still change?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "March 2001: My First Six Months".
2. Staats, *The WoW Diary*, section "July 2001: Nine Months Down the Tubes".
3. Staats, *The WoW Diary*, section "July 2001: Nine Months Down the Tubes".
4. Staats, *The WoW Diary*, section "September 2001: Belated Progress with Dungeons".
5. Staats, *The WoW Diary*, section "March 2001: My First Six Months".
6. Staats, *The WoW Diary*, section "March 2001: My First Six Months".
7. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V5 at 0:11, 0:59 and 1:30.

8. Candlesan, commentary, V5 at 2:15.
9. Candlesan, commentary, V5 at 2:48.
10. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
11. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".
12. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
13. Staats, *The WoW Diary*, section "Dungeons: The Last Hurdle".
14. Staats, *The WoW Diary*, section "July 2001: Nine Months Down the Tubes".
15. Staats, *The WoW Diary*, section "August 2002: Ingenuity with Cheats and Bugs".
16. Staats, *The WoW Diary*, section "E3 2003".
17. Staats, *The WoW Diary*, section "April 2003: A Slightly Higher Profile".
18. Mark Rosewater. Kaladesh Ingredients, Part 1. Making Magic, `magic.wizards.com`, 2016, section "Building Up Energy".
19. Rosewater, "Kaladesh Ingredients, Part 1", section "Energy Conservation".
20. Mark Rosewater. Nuts & Bolts 7: The Three Stages of Design. Making Magic, `magic.wizards.com`, 2015, section "The Integration Stage".
21. Staats, *The WoW Diary*, section "July 2003: Unexpected Giants".
22. Mark Rosewater. Looking Back, Part 2. Making Magic, `magic.wizards.com`, 2024, section "2013 Best Addition (Devotion)".
23. Rosewater, "The Three Stages of Design", section "The Vision Stage, Iteration".
24. Mark Rosewater. Nuts & Bolts 18: Layering Your Mechanics. Making Magic, `magic.wizards.com`, 2026, paragraph opening "For early playtests, make 40-card decks".
25. Rosewater, "Layering Your Mechanics", paragraph opening "At this point, it's possible your first mechanic".
26. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, section "R&D is only human".
27. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "A Complex Problem".
28. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section "Kaladesh, Lessons".
29. Staats, *The WoW Diary*, section "July 2003: Unexpected Giants".
30. Staats, *The WoW Diary*, section "June 2004: Public Beta 2.0".
31. Staats, *The WoW Diary*, section "July 2001: Nine Months Down the Tubes".
32. Staats, *The WoW Diary*, section "June 2003: A Crunchier Crunch".

Chapter 24

Technical Constraints Are Game Mechanics Too

When John Staats joined the *World of Warcraft* team, two colleagues explained to him how the game might work. Like *Diablo*, it would let players adventure alone while disconnected, which would save the company bandwidth, and connect them to the servers only when they wanted to group up. John Cash, the lead programmer, stood behind them shaking his head. Information about weapons and monsters could not be kept on a player's computer, he said, because anything kept there can be hacked. "Nothing valuable can ever be on the client." A producer closed the matter on the spot: then they would not do that.[1]

Nobody in that conversation was designing a mechanic. They were deciding where data would sit. Offline play died in one sentence, and with it every design in which a player's progress belonged to the player's own machine. From then on every character, every item and every monster's decision lived on a server, and each player saw the world through a copy that the server kept updating.

The decision cost nothing to state and bound every idea that came after it. A whole category of game was removed before anyone had drawn it, by a programmer thinking about cheats.

24.1 Where the Data Sits Decides What Can Be Played

Staats later states the rule in full. Nothing on the server could be expensive to compute, because one server computes for everybody. Nothing on the client could be critical, because the client belongs to the player. So the server kept the cheap and valuable things, such as inventories and the behavior of monsters, and the client did the expensive work whose hacking would

break nothing, such as drawing the art and handling collision. Game ideas that did not fit were rejected.[2] The first casualty was precise hit detection, too costly for a crowded server and impossible to trust to a client.[3] The engine ran a huge open world full of players well and would have made a poor shooter, because it was never built for fast action or for synchronizing minute actions.[4]

Money pushed the same way. Late in 2003 the game sent 1.7 kilobytes a second, less than a tenth of what other games of its kind used, and bandwidth was potentially the team's largest running cost after customer service.[5]

The server also reasons about a poorer world than the one on screen. The engine handles three sorts of geometry: what is drawn, what blocks movement, and where monsters may walk. Only the third lives on the server, and in it stairs are flattened to a plane and doors do not exist, which is why monsters can run through closed ones.[6] Chapter 3 tells what players did with the gaps in that floor plan. A mechanic has to make sense in the server's world, not the artist's.

Where the world had seams, design hid them. In January 2002 Joe Rumsey showed that a player could cross from one server's territory into another's without noticing, by fighting and looting a monster across the invisible line, and the team stopped having to guess how to carve the world into separate servers.[7] A dungeon stayed apart: a private copy on its own server, joined to the zone only at the entrance. The Deadmines was to end with the reveal of a warship moored offshore, and a player who killed the last boss and dived in would swim toward a shore that, for that server, did not exist. After days of discussion the builder put the ship inside a giant cavern.[8] The sea is flat for a related reason. Chapter 8 tells how lag ruled out three-dimensional waves; the transitions between rivers went behind waterfalls, which turned out to be picturesque.[9]

And the copies of the world disagree. Every client reports its position to the server and the server broadcasts everybody's position back,[10] so what a player sees of the others is always a little old, by an amount nobody can see. Chapter 10 tells how that delay wrecked Gruul's scramble and how the repair became a boss ability, a slow that gives the clients time to agree. Chapter 8 tells the same of spell batching, a server habit of processing messages in groups that *WoW Classic* rebuilt by hand because combat had been balanced on top of it.[11]

Costikyan's catalog of uncertainty, set out in Chapter 5, lists the kinds a designer puts into a game, among them performative uncertainty, whether

a player can execute, and hidden information, what a player may not see.[12] Latency adds a kind nobody chose and no player can answer by playing better. A cave, a waterfall and a slow are three responses to limits of this sort, and together they give the principle.

Hardware, network, interface and tools decide which mechanics can exist. At that boundary the designer's work is to find where it really is, to remove the uncertainty a player cannot act on, and to turn what the boundary forces into play.

The first of those jobs is the hardest, because the person who reports a limit is rarely the person who has measured it.

24.2 A Door Is as High as the Camera Allows

The camera follows the character from behind, and indoors it collided with walls and ceilings. When it slid behind a wall hanging, the screen filled with the back of a picture. In November 2001 the team worried that so much camera movement would disorient players, and Staats did not yet know which would give, the camera or the buildings.[13] Months earlier the level designers had been unable to learn whether the camera would bump the tops of doorways and how high ceilings had to be.[14] A lens was setting the height of every door in the game, and with it how much of a fight indoors a player could see.

Other limits were smaller than claimed. The programmers had refused transparent water. Spell effects are drawn with transparent textures, and when one transparent surface sits in front of another the graphics card can draw them in the wrong order, so a spell cast above the water would appear to be beneath it. They had no time to fix it. Staats took a prop with a transparent texture, scaled it up and used it as a mock waterline. The glitch showed only when he channeled a spell from the shore into the water, and it was "utterly unnoticeable in the heat of combat." A producer promised transparent water. Staats grants that the programmers were right that the problem existed. His test showed its size.[15]

The same months record tents placed on open terrain against a pathing warning that proved milder than feared, and a thorn canopy that broke the engine's size limit until, after weeks of vetoes, the programmer wrote an exception.[16] Chapter 4 tells the sequel to the Deadmines cavern: a longer draw distance, bigger rooms everywhere, and a cost in frame rate that testing found negligible.

Tools draw the boundary as surely as hardware. Chapter 23 tells how the first dungeon editor made finished work nearly impossible to resize, and Chapter 22 how combat stayed fake until designers had an editor for spells. What a tool makes cheap gets tried. What it makes expensive gets argued about instead.

24.3 What a Limit Costs

Chapter 4 called the actions open to a player the feasible set. The designer has one too: every mechanic the platform can deliver. Choosing inside it is constrained optimization, and three of its ideas sort the cases above without any algebra.

A constraint either binds or does not. A binding constraint is one the chooser is pressed against: remove it and a better choice opens. The line between client and server bound. Offline play, precise hit detection and a moving sea lay beyond it, and the team gave up each of them.

A binding constraint has a ***shadow price***: what the chooser would gain if the limit were loosened a little. The price says whether a limit is worth fighting. The delay at Gruul was dear, since without an answer the scramble the fight existed for was a lottery, and it was paid for in design, with five seconds of slow.

A limit believed to bind may not. A constraint with slack costs nothing to obey and returns nothing when attacked, and one nobody has measured could be either kind. Transparent water and the draw distance were limits priced near zero, and the only way anyone learned it was by building past them. The mock waterline is a shadow price found by experiment, with a prop.

24.4 The Card With No Back

Magic: The Gathering has no server. Its platform is a printed card, and the card's limits are as firm as a modem's. For *Innistrad* the designers printed werewolves on both faces of the card, and that answer broke something no rule had needed to address (Chapter 7 reads the break as a problem of hidden information). The rules sort the game's zones into public and hidden, and a player's library and hand are hidden.[17] What keeps them hidden is physical. Every card ever printed had the same back, so a face-down card is anonymous.[18] A card with a werewolf where the back should be announces itself from the top of a deck.

The first plan was a pair: an ordinary card in the deck that fetched its two-faced partner from outside the game. It died at the printer. Production could promise, at considerable expense, to match the pairs about 90 percent of the time, which would leave roughly one pack in ten holding half a pair. The shipped answer had two parts. Opaque sleeves cover the back, and about 90 percent of players already sleeved their constructed decks. For play straight from fresh packs there was a checklist card, a single universal card with a normal back, marked to show which double-faced card it stands for, that takes that card's place wherever identity must stay secret. The problem had become one of logistics and not of game play.[19]

The logistics were exact. A checklist is marked with the one card it represents, the mark must not show through the back, and a deck that uses a checklist for one double-faced card must use them for all.[20] The pack paid as well: the checklist took the place of a basic land in roughly three packs out of four, and the set's basic lands went from four pieces of art to three.[21] Part of R&D wanted the cards out, and critics held that the common back was a rule never to break.[22][23] Players made them a great success.[24]

Paper's second limit sits in the players. No engine tracks the state of a game of *Magic* or reminds anyone that a condition has been met. Energy was reworked for the tracking it demanded (Chapter 23),[25] and New World Order cleared small tracked effects out of the commonest cards (Chapter 21).[26] Revolt went into the record as a mistake, first of all because its condition was too easy to miss.[27] In a paper game the players are the server, their memory is its capacity, and a trigger nobody notices is a message that never arrived.

The turn itself is partly bookkeeping for a table with no engine. A keyword for creatures that attack without being turned sideways cannot exist on paper without a fixed moment for attacking, since nothing on the table would show whether the creature had already attacked; a digital game can remember that for the players. A digital client that stops at every point where a response is allowed makes visible windows that players knew of and had never really seen.[28] The turn structure is a workaround for the platform, and a different platform teaches a rule that paper leaves hidden.

24.5 When the Limit Is a Guess, and When It Has Moved

Turning a limit into play is the wrong move when the limit is a guess. Had the dungeon builders designed gracefully around opaque water, the

submerged temples Staats had built would have stayed invisible, and the accommodation would have been praised as craft. Testing has its own bill. Level designers serve so many masters that they sometimes forget their limitations, Staats writes, "robbing Peter to pay Paul," and props overused in outdoor zones did degrade the engine's performance, until programmers took the disputes to the producers.[29] Some limits are real, and a team that attacks every one of them pays for the ones that hold.

Constraints also move, and the workaround stays behind. Gruul's slow answered the connections of 2007, when many players were still on dial-up.[30] Batching was kept after the modern server had left it behind. The interface shows the pattern at the largest scale.

Second-Order Effect: Open the interface, inherit the players' code

The interface was a limit of a different kind: the team could build only so much of it. When developers tested a raid in October 2004 there was no raid interface at all, and the fights were measured with damage meters that were beta add-ons built by fans, because the team had no internal tools for the job.[31] The gap had been filled by a decision to let players script the interface, and Chapter 15 tells what they wrote in the first week.[32] The workaround was generous and it worked. It also made the players' code part of the platform, and about two decades later the studio was restricting the combat data add-ons may read.[33] A limit on the team's time had been traded for a dependency on its audience's tools.

A workaround that players have built habits on is no longer a workaround. Chapter 25 takes up what a designer owes to decisions of that age.

The Designer's Test

1. What does the platform do to the information each player holds at the moment of decision?
2. Which uncertainty here can the player answer, and which is noise?
3. Who told you this limit exists, and has anyone priced it with a test?
4. Who or what is keeping track of state, and what happens when they miss?

5. Which of today's workarounds will still be in the game when the constraint is gone?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "March 2001: My First Six Months".
2. Staats, *The WoW Diary*, section "Gameplay".
3. Staats, *The WoW Diary*, section "Why MMOs Are So Difficult to Create".
4. Staats, *The WoW Diary*, section "Gameplay".
5. Staats, *The WoW Diary*, section "December 2003: Stepping on Toes".
6. Staats, *The WoW Diary*, section "Programming: The First Hurdle".
7. Staats, *The WoW Diary*, section "January 2002: The Stitches of a Seamless World".
8. Staats, *The WoW Diary*, section "April 2002: The Occasional Paradox".
9. Staats, *The WoW Diary*, section "November 2001: Client-Server Headaches".
10. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V2 at 1:17.
11. Kaivax. Spell Batching in Classic. WoW Classic Bug Report forum, `us.forums.blizzard.com`, 2019.
12. Greg Costikyan. *Uncertainty in Games*. The MIT Press (Playful Thinking series), 2015.
13. Staats, *The WoW Diary*, section "November 2001: Client-Server Headaches".
14. Staats, *The WoW Diary*, section "March 2001: My First Six Months".
15. Staats, *The WoW Diary*, section "December 2003: Stepping on Toes".
16. Staats, *The WoW Diary*, section "December 2003: Stepping on Toes".
17. Wizards of the Coast. Magic: The Gathering Comprehensive Rules (effective January 16, 2026). `media.wizards.com`, 2026, rule 400.2.
18. Wizards of the Coast. Double-Faced Card Rules. *`magic.wizards.com`*, 2011, section "Putting Double-Faced Cards in Your Deck".
19. Mark Rosewater. Every Two Sides Has a Story. *`magic.wizards.com` (read from Wayback copy)*, 2011, sections "Double-Face the Music" and "Double-Face Off".
20. Wizards of the Coast, sections "Putting Double-Faced Cards in Your Deck" and "Double-Faced Cards in General".
21. Rosewater, sections "Double-Face the Music" and "Double-Face Off".
22. Mark Rosewater. Innistrad Design Handoff Document. Making Magic, `magic.wizards.com`, 2021, section "Transformation".
23. Rosewater, section "Double-Face Off".
24. Mark Rosewater. Looking Back, Part 2. Making Magic, `magic.wizards.com`, 2024, section "2011 Best Addition (DFCs)".
25. Mark Rosewater. Kaladesh Ingredients, Part 1. Making Magic, `magic.wizards.com`, 2016, sections "Building Up Energy" and "Energy Conservation".
26. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "A Whole New World".
27. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, section "Kaladesh, Lessons".
28. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 4 at 7:46 and 8:15; episode 22, part A, at 19:56.

29. Staats, *The WoW Diary*, section “December 2003: Stepping on Toes”.
30. Candlesan, commentary, V2 at 1:28.
31. Staats, *The WoW Diary*, section “October 2004: World’s End”.
32. Staats, *The WoW Diary*, section “March 2004: Public Beta 1.0”.
33. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, section “ADDONS & THE TRICKLE DOWN EFFECT”.

Chapter 25

Designing a Game That Will Change

At the E3 trade show of 2003, John Staats spent three days at the Blizzard booth explaining *World of Warcraft* to anyone who would listen. Fans jumped up and down. Distributors and executives did not, so he tried facts: six languages, and partners around the world lining up to support it. Then he offered what he calls a personal suspicion, that the game could last for twenty years and that the team had a great many ideas it had not yet built. "People searched my face for signs that I was kidding them," he writes. He was not.[1]

One visitor had more reason than most to believe him. Staats gave a demonstration to Richard Garfield, the designer of *Magic: The Gathering*, and explained how trade skills worked and how their rewards tied into equipment. Garfield was cordial, and "his countenance was unreadable."[2] His own game had premiered in the summer of 1993.[3] By the time he stood at that booth it had been adding cards to itself for a decade.

Staats was selling a product that would never be finished, and offering that as its strength. A year later his team was living inside the description.

25.1 A Launch Is a Point on a Line

In May 2004 Mark Kern had to repeat to the team that combat between players was being postponed, not cut, and that this was a live game, so a missing feature or a bug was not the end of the world. Some employees openly declined longer hours because the game would never be finished. Everyone knew about the expansions, continents and patches ahead: "it seemed as if shipping the game wouldn't provide the team with any closure."[4] When the master disc left for the factory in October, the wish list ran from combat between players to housing, expansion zones and new dungeons, and four dungeons were assigned to the second patch.[5]

A boxed game is judged once. Its rules meet each other on the day it ships, and whatever they add up to is the game. In these two games every patch, expansion and set is added to everything already in play, and the players keep their characters, their cards and what they have learned. The designer of the tenth addition works inside the decisions of the first nine. *Magic* is the purer case, because its past is printed: a card in a player's box keeps the text on it, and each year's new cards, mechanics and concepts pile on top of the old ones.[6] *World of Warcraft* can rewrite any number in a patch. It cannot rewrite what its players remember owning.

So a rule in these games has two lives. In the first it produces play, this season, among the rules that exist. In the second it sits in the foundation while other rules are built on it, and it decides which of them can be built. Call that second quantity ***design space***: the set of things a designer can still make that will work beside what already exists. A release spends some of it and, if it is well made, opens some.

The natural policy ignores the second life. Make each release as good as it can be, add what excites now, fix what is weak now. Players push for exactly that. They expect new abilities each time the level cap rises, and designers feel pressure to repair a neglected ability when cutting it might be wiser.[7] A publisher faces the commercial form of the same pull: nobody buys the next small expansion if none of its cards is good enough for tournaments.[8]

25.2 Good Releases Add Up to a System Nobody Chose

The additions do not stay separate. *Magic*'s complexity crept until fewer people were learning the game,[9] and *World of Warcraft* added abilities faster than it removed them until some players felt they needed dozens of keybinds;[10] Chapter 21 tells both as budgets. Power runs the same route with a ratchet, the sum Chapter 19 does: each set must outdo its predecessors,[11] and the strongest card of this year becomes next year's floor, until raising every set breaks the game.[12] No single release did any of this. Each passed its own review; none was reviewed against the total.

The third accumulation needs no additions at all. When *Cataclysm* rebuilt the original low-level zones, the reason given was that old content had come to feel old. Westfall had reached seven years of age with trees and quests of the same vintage, and the order in which classes learned their spells had aged poorly too.[13] Every new zone had moved the standard, and the part every new player meets first fell behind by standing still.

Players outrun every forecast. In January 2004 the launch plan could carry 30 percent of the user base online at once, double the 15 percent that was the highest any game of the kind had reached. The game launched at 90 percent, six times the old record, and never fell below 50.[14] *Magic*'s first rulebook asked only for a deck of at least 40 cards and set no limit on copies, because Garfield assumed a player would spend on it roughly what people spend on a board game, so a small collection would rarely hold more than one or two of the most disruptive rares. Players bought boxes of boosters. Garfield had seen the possibility and filed it as a problem to solve if the game succeeded wildly.[15] Both forecasts erred toward success, and both games absorbed the difference after release.

And players treat what they hold as theirs. When the company reprinted scarce early cards for people who had missed them, those who already owned the cards were unhappy, and the effect on their value had not been thought through.[16] An expectation that earlier releases had created was touched, and an expectation behaves like a rule the designer never wrote and cannot easily repeal.

Second-Order Effect: Postponed is not free

A live game turns a cut into a postponement, and that looks like a pure gain. Two mid-level dungeons, Maraudon and Dire Maul, were lagging before launch and were pushed to a later patch. The delay cost twice. It weakened the team's ability to learn how players looked for groups, valued boss loot and behaved in dungeons, because testers who leveled past the hole found high-level dungeons they could not yet run. And when the two dungeons were finally released, they were too low-level for much of the audience.[17] The option to build them later was kept. The players they were built for had moved on, and so had the months in which they would have taught the team something.

Judge a rule twice: by the play it produces now, and by what it leaves possible later.

The working idea is ***option preservation***. A rule preserves options when designs that do not exist yet can be placed beside it without breaking, and destroys them when every later design has to be checked against it, weakened to fit it or abandoned because of it. *Magic*'s once-per-turn clauses are the card-sized case. Restrictions are part of a design and not a by-product of it, because controlling when and how often an effect happens

controls how large it may be.[18] In a game that never stops adding to itself, a clause that caps an effect today also caps what any future card can do with it. A playtest runs the first test. The second has to be reasoned about, because what the rule must coexist with does not exist yet.

25.3 Rotation, Bans and a League a Year Ahead

Magic met the problem first and built the oldest instruments for it. The 60-card minimum and the four-copy limit were both added after release, once tournaments showed what the original rules allowed.[19] They did not solve power. The strongest early cards were too strong for new cards to compete with,[20] and something had to keep the power level from spiraling. The answer was a second format restricted to the last few years of sets, now called Standard. Players met it with the largest outcry of any announcement in memory, asking whether the cards they owned were about to become obsolete, and it went on to become the favorite format by a wide margin.[21] The movement of old sets out of it is called rotation.

Rotation is option preservation built into the structure of the game. In Standard every card has a date on which it leaves, so every mistake has one too, and next year's designer must coexist with a few years of cards instead of every card ever printed. In 2023 a set's stay was lengthened from two years to three, for the longevity of the cards players own and room to build on mechanics over time, in a year when regular Standard play had dried up in many stores while other formats grew.[22] A short window protects tomorrow's designs from today's errors. A long one protects today's purchases from tomorrow's sets. No setting does both.

Rotation does not cover a problem whose cause is not a card. *Mirrodin* loosened the color requirement on artifacts, the giving way Chapter 4 tells.[23] It was released on October 2, 2003, and *Kaladesh*, a later artifact block, became legal on September 30, 2016, just under thirteen years later.[24] The artifact problem bit the team again and had to be solved before the next artifact block.[25] A missing constraint does not rotate.

A ban is what repair looks like when the past is printed. The January 15, 2018 announcement removed two cards from Standard's energy decks, whose figures Chapter 3 gives, and one of its stated grounds was that rotation had not meaningfully weakened the deck.[26] The instrument built to let mistakes expire had come and gone, and the mistake remained. Before banning Aetherworks Marvel in June 2017, R&D considered adding answers

from outside the format, restricting the card and rewriting it by errata, and rejected each as creating more problems than it solved.[27]

The better instruments act before release. Proliferate adds to whatever counters are already in play, so it adapts to each set it meets, and its designers called it a shift in how they thought about backwards compatibility, which is hard to achieve many years into a game.[28][29] *Fifth Dawn* had to turn sharply inside a block plan already fixed, and the lesson drawn was to plan ahead.[30] Planning, once worked out together with a block's first set, was moved ahead of design.[31] Testing moved the same way. R&D ran an internal league that played the game as it would exist six months ahead, found six months enough to identify problems and not enough to fix them, and pushed the Future Future League out to a year.[32]

25.4 What Players Have Built on It

World of Warcraft's largest addition is a class, and the death knight left a record to design against: made too strong so that a new class would not feel neglected, then brought back into line, it left players feeling the rug had been pulled from under them (Chapter 20). The monk was planned with a simpler resource model and fewer swings.[33] A correction that would have been a line in a spreadsheet before release had become, after it, a withdrawal from people who had built on the number.

Numbers carry the same debt. Chapter 19 follows the squish that kept every ratio and still had to promise it was not a nerf,[34] and the 2020 change that turned level 120 into level 50.[35] Each number had been raised by a release that was better for raising it. A growth rate that feels generous in one expansion is a commitment to compress the whole game later. The split between the two factions, told in Chapter 4, is a standing order for content of the same kind, payable in every release for as long as the wall stands.

Some of the future is written by players. The scriptable interface of Chapter 24 ended, in the contest Chapter 15 follows, with the combat data add-ons may read restricted, classes simplified in compensation and official tools built for the most necessary jobs. The encounter designer's verdict was that some mechanics in the opening raid tier are possible now that might not have been in the recent past.[36] An open interface had been closing a region of design space, the puzzle a group solves by talking, and restricting it gave the region back at the price of tools players had built on. *Magic* has its counterpart. A card's usefulness shifts with the field of decks around it,[37] and Commander, a format players adopted, gives each player 40 life and

several opponents and so undid strengths white and red had been designed around. The game has to evolve as players adopt formats (Chapter 17).[38]

Two tools price all this. The first is ***backward compatibility***: the cost of a new rule rises with the size of the catalog it must coexist with, because every existing piece is a possible interaction. Rotation attacks the catalog. Proliferate attacks the interaction, by being written against a category and not a list. The second is accumulation: a cost too small to matter in one release, paid in every release, becomes the main cost.

Design equation (a conceptual framework, not a law)

worth of a rule over the life of the game = play it produces now + design space it leaves open − what every later release must carry because of it

The natural policy measures the first term only; the other two are invisible on release day. The player feels them as what Costikyan calls development anticipation,[39] uncertainty about what the game itself will become: whether the class just chosen will be reduced, whether the cards just bought will rotate.

25.5 Insurance Has a Premium

Keeping an option open has a carrying cost, and today's player pays it. Some of the newer restrictions feel bad to read, a cost Chapter 4 sets out,[40] and rotation is the same payment at a larger size: the device that keeps power from spiraling tells a player the deck has an expiry date. The 2023 change set that price lower.

Insurance bought against the last problem can cause the next. The 2018 announcement also banned Rampaging Ferocidon, designed when a different deck dominated; in the format that arrived, it shut down the ways other decks could fight the strongest red deck.[41] *Amonkhet* was designed to swing away from *Kaladesh*'s tight synergy, and the pendulum swung too far.[42] The principle asks a designer to reason about releases that do not exist, and the only evidence is the release just shipped. A game that changes steers by its own wake.

Refusing to change has a bill too. A closed card pool shields its players from rising power and keeps a deck bought once valid, but if one deck comes to dominate it leaves a single lever, the ban, and each ban takes away

something a player paid for.[43] A game that declines to change gives up every instrument in this chapter except the bluntest.

The Designer's Test

1. What will this rule have to coexist with in five years that does not exist yet?
2. What does this release add that every later release will have to carry?
3. If this has to be reduced later, what will players have built on it by then?
4. What are you postponing, and who will still be there when it arrives?
5. Which changes to this game will come from the players and not from you?

Notes

1. John Staats. *The WoW Diary: A Journal of Computer Game Development*. whenitsready, 2019, section "E3 2003".
2. Staats, *The WoW Diary*, section "E3 2003".
3. Mark Rosewater. Eternal Formats of the Spotless Mind. Making Magic, `magic.wizards.com`, 2016, section "A Long Time Ago".
4. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
5. Staats, *The WoW Diary*, section "October 2004: World's End".
6. Mark Rosewater. New World Order. Making Magic, `magic.wizards.com`, 2011, section "Problem Solving".
7. Greg Street. Dev Watercooler: Number of Abilities. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "The Human Factor".
8. Mark Rosewater. When Cards Go Bad. Making Magic, `magic.wizards.com`, 2002, section 1.
9. Rosewater, "New World Order", section "Problem Solving".
10. Blizzard Entertainment. Dev Watercooler: Pruning the Garden of War. `worldofwarcraft.blizzard.com`, n.d., section "Ability Pruning".
11. Mark Rosewater. In the Mailbag. *Making Magic, `magic.wizards.com`*, 2008, reply beginning "You caught us".
12. Rosewater, "When Cards Go Bad", section 1.
13. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., answer on main goals.
14. Staats, *The WoW Diary*, section "January 2004: One Year Left".
15. Rosewater, section "A Long Time Ago".

16. Mark Rosewater. Twenty Things That Were Going To Kill Magic. *Making Magic,* *magic.wizards.com*, 2013, items 3 and 5.
17. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
18. Mark Rosewater. Why So Many Restrictions? Making Magic, magic.wizards.com, 2024, section "Why even have restrictions?".
19. Rosewater, section "A Long Time Ago".
20. Rosewater, paragraph beginning "Which brings us to 1995".
21. Rosewater, item 3.
22. Aaron Forsythe and Billy Jensen. Revitalizing Standard. *Announcements,* *magic.wizards.com*, 2023, paragraphs beginning "That step is" and "Our intent with this change".
23. Mark Rosewater. Lessons Learned, Part 1. Making Magic, magic.wizards.com, 2023, section "Mirrodin".
24. Wizards of the Coast. Set release dates: Mirrodin, Time Spiral, Lorwyn, Morningtide, Kaladesh. Gatherer set pages and Kaladesh Release Notes, wizards.com, n.d., Mirrodin set page and Kaladesh Release Notes.
25. Mark Rosewater. State of Design 2017. Making Magic, magic.wizards.com, 2017, sections "Kaladesh, Lessons" and "Wrapping Things Up".
26. Ian Duke. January 15, 2018 Banned and Restricted Announcement. *magic.wizards.com (read from Wayback copy)*, 2018, section "Temur Energy and Other Energy Variants".
27. Aaron Forsythe. June 13, 2017 Banned and Restricted Announcement. *magic.wizards.com (read from Wayback copy)*, 2017, paragraph beginning "To that end, we considered".
28. Mark Rosewater. Looking Back, Part 2. Making Magic, magic.wizards.com, 2024, section "2010 Best Addition (Proliferate)".
29. A counter is a marker placed on a card or a player to track a lasting change, such as a creature made permanently larger.
30. Rosewater, "Lessons Learned, Part 1", section "Fifth Dawn".
31. Mark Rosewater. Advanced Planning. Making Magic, magic.wizards.com, 2014, sections "Before:" and "After:".
32. Mark Rosewater. New New World Order. *Making Magic,* *magic.wizards.com*, 2013, opening paragraphs.
33. Blizzard Entertainment. Blizzard Insider 42: Meet the Pandaren Monk (interview with Greg Street). worldofwarcraft.blizzard.com, n.d., answers on adding a class and on death knight lessons.
34. Blizzard, "Pruning the Garden of War", section "Stat Squish".
35. Blizzard Entertainment. Explore the Updates Arriving with the Shadowlands Pre-Expansion Patch. *Blizzard News,* *news.blizzard.com*, 2020, section "Levels, Items, and Stats Adjusted".
36. Jaydaa. Challenge with Communication and Coordination, Not More Bullets: Midnight Interview with PC Gamer (quoting Dylan Barker). Wowhead, 2026, sections "ADDONS & THE TRICKLE DOWN EFFECT" and "OPENING UP ENCOUNTER DESIGN".
37. Rosewater, "When Cards Go Bad", section 4.
38. Rosewater, "Why So Many Restrictions?", section "Okay, Richard built plenty of restrictions into the game".
39. Greg Costikyan. *Uncertainty in Games.* The MIT Press (Playful Thinking series), 2015.
40. Rosewater, "Why So Many Restrictions?", section "Does it bother you that some of the new restrictions feel bad to read?".
41. Duke, section "Ramunap Red".
42. Rosewater, "State of Design 2017", section "Amonkhet, Lessons".

43. Distraction Makers. Design discussion videos. YouTube, `youtube.com/@DistractionMakers`, episodes listed in Video Sources, 2023–2026, episode 17, part A, at 4:35, 8:19 and 9:35.

The Designer's Real Job

Before Loatheb had a debuff, he had a roster problem. The first rule his designers considered was to switch healing off for the fight, and they dropped it because it would get healers benched.[1] Nobody in that moment is casting anything. The person being reasoned about is a raid leader, some evening before the encounter, deciding whom to invite. Every later version was judged the same way. The penalty shared on the tank paid for more heals, not fewer, and healers who knew the plan abandoned it under load; the rule that shipped, one heal per healer per minute, removed the habitual answer and left more than one standing.[2] Then the fix sent its bill elsewhere. Less healing meant tanks short of rage, so the room gained spores that let players deal damage without generating threat,[3] and a fight with so little healing leaned on potions that developers could conjure with a command and players could not. The requirement shipped after an internal debate.[4] Only after the boss was finished did his designer learn that Loatheb is an anagram of healbot, which was not intentional and does seem appropriate.[5] That is the book in one fight: a designer choosing a debuff was in fact building the incentives for one human decision under pressure, stop healing and attack, and then dealing with the other decisions the incentives produced unasked.

The Same Order of Events

The cases in this book have almost nothing in common as mechanics. A loot rule, a random throw, a health bar, a drop table, a card back and an artifact's color requirement belong to different games, different decades and different departments. What repeats is the order of events. Someone typed a rule with a behavior in mind. The rule paid for something, showed something, removed something or promised something. A person on the other side read that, and not the intention, and chose. Enough people chose the same way that the choice became the game.

The first internal alpha of *World of Warcraft* let anyone loot any corpse, and a game being built around other people paid best, for a week, to the melee players who left the crowd to play alone (Chapter 1).[6] Nobody wrote solitude into the rule. The rule paid for it, and players collected.

Gruul holds the cleanest pair of uncertainties in the book (Chapter 10). Ground Slam throws every raider a random distance in a random direction, uncertainty the designer wanted because it keeps a raid from walking to assigned spots.[7] The disagreement between two screens about where a player stood was uncertainty the network supplied, and the designer spent a boss ability on it, a slow that gives the clients time to synchronize.[8] A player answers the first by looking and moving. No skill answers the second. A designer who cannot tell those two kinds apart is grading connections and calling it play.

Skeram's copies were meant to be told apart by watching health bars. Players sent a hunter's pet, which stayed on the real one, and after all the disguising the fight still had an instant tell (Chapter 15).[9] The system displayed a fact its author had not chosen to display, and players read the system, not the author.

Three ordinary swords added to a flat loot table lower the legendary's drop rate, a probability that moved when unrelated content changed and that nobody decided to move (Chapter 18).[10] The ledger has the same structure as the fight: a rule's effect lives in what surrounds it.

And the platform writes rules too. A lead programmer's one sentence, that nothing valuable can ever be on the client, removed offline play before anyone had designed it (Chapter 24),[11] and a card printed on both faces turned a printing problem into a checklist card and a lost land slot.[12] A constraint *Magic* loosened on artifacts was still biting when a block returned to the theme thirteen years later, because a decision in a game that keeps adding to itself has to survive designs that do not exist yet (Chapter 25).[13]

Where the record holds a second attempt that worked, the designers got there by asking what a specific person will do with this, not by adjusting the size of the first attempt.

A Fix Moves the Cost

Read side by side, the limits of every chapter make one record, and it is the book's hardest finding. A fix moves the cost. It rarely removes it.

Pruning the talents nobody was really choosing drew a split verdict: the specialization choice it moved to level 10 was acclaimed, and the rest left the

team unhappy, as content lost for little gain (Chapter 2).[14] The cascading roll fixed the rate of a tier and passed the dilution down to the items inside it (Chapter 18). Rotation protects tomorrow's cards from today's mistakes at the expense of the deck a player already owns (Chapter 25). Private auras closed a leak, and the people who built them judged the cure sometimes worse than the problem (Chapter 15).[15]

The cost usually lands on someone other than the person the fix was for. It moved from healers to tanks, from experts to newcomers, from an organized guild to strangers in a pick-up group, from this year's set to a later one, and from the player to the designer's own schedule. A repair judged only from the chair it was built for will look free. Who pays, besides the people who benefit, sorts more outcomes than any label does.

The principles also break against each other. Chapter 10 says to keep an answer from being worked out in advance, and the game's current encounter leads keep fully plannable fights on purpose, because planning is a pleasure some raiders came for.[16] Chapter 16 says to doubt the players' remedy, and the auction house followed because players built one first.[17] Chapter 22 says to name the system before building its parts, and the diarist it relies on argues for building from moment-to-moment play. None of these is a contradiction for a better principle to resolve. Each marks a place where two things a game wants cannot both be had in full, and the designer decides which one this fight, this set or this audience is for. The principles locate the decision. They do not make it.

The Rule Is the Interface

A player meets a rule: a slow after a throw, a card that stands in for another, a format that retires a deck. It is the entire visible surface of a long chain of other people's reasoning, and nothing on that surface shows who was expected to panic, who would be left out and who would pay. Players do not need any of it. They need only to do what the rule makes sensible, and when the design is right, what is sensible is what the designer hoped for. The rule is where the designer's hands and the player's hands touch the same object. It is the interface, not the thing being made.

The thing being made is the behavior that emerges from it. Game design is the architecture of choices, constraints, information, uncertainty, rewards, social interactions and expectations, and never of rules alone. A designer who works on the rule and inspects only the rule is polishing the interface. The work is on the other side of it, in a room the designer will not be in,

among people who are frightened, hurried or simply trying to win, and who will read what was built more closely than its author did.

Nobody foresees that room exactly. When the developers testing the first raid kept losing half their number to its first boss, Jeff Kaplan told them, "Trust me, once the raiders figure this out, it will be trivial."[18] The same diary preserves a line from the design team's whiteboard: "Vision is what some people claim to have when they've guessed correctly."[19] The evidence here is the designers' own record of two games, and it shows what the mechanism looked like each time someone wrote it down, not how often reasoning this way is right. The thesis does not need more. Good designers do not foresee behavior. Behavior is what they are trying to foresee, and the ones in this record found out they were wrong by watching it.

A designer writes rules. The rules create incentives, whether or not anyone intended them. People facing those incentives make decisions. Decisions made by many people, many times, become behavior. Behavior in a system with enough moving parts becomes outcomes that nobody wrote down. Only the first link can be typed. The designer's real job is the other four (Figure 25.1).

Notes

1. Candlesan. Encounter and systems design commentary. YouTube, `youtube.com/@candlesan`, segments V1–V7 listed in Video Sources, 2026, V6 at 0:12.
2. Candlesan, commentary, V6 at 1:25.
3. Candlesan, commentary, V6 at 1:43.
4. Candlesan, commentary, V6 at 2:26.
5. Candlesan, commentary, V6 at 2:35.
6. John Staats. *The WoW Diary: A Journal of Computer Game Development.* whenitsready, 2019, section "September 2002: Internal Alpha 1.0".
7. Candlesan, commentary, V2 at 0:59.
8. Candlesan, commentary, V2 at 2:00.
9. Candlesan, commentary, V3 at 2:47.
10. Candlesan, commentary, V1 at 0:50.
11. Staats, *The WoW Diary*, section "March 2001: My First Six Months".
12. Mark Rosewater. Every Two Sides Has a Story. *`magic.wizards.com` (read from Wayback copy)*, 2011, sections "Double-Face the Music" and "Double-Face Off".
13. Mark Rosewater. State of Design 2017. Making Magic, `magic.wizards.com`, 2017, sections "Kaladesh, Lessons" and "Wrapping Things Up".
14. Greg Street. Cataclysm Post Mortem. Blizzard Entertainment, `worldofwarcraft.blizzard.com`, n.d., section "What didn't work out as planned?".
15. Drew De Sousa and Taylor Sanders. The Future of Raid Encounter Design: Part 2 of Wowhead's Exclusive 11.2 Interview. Wowhead, 2025, answer beginning "That's awesome. I know you guys have been responding".

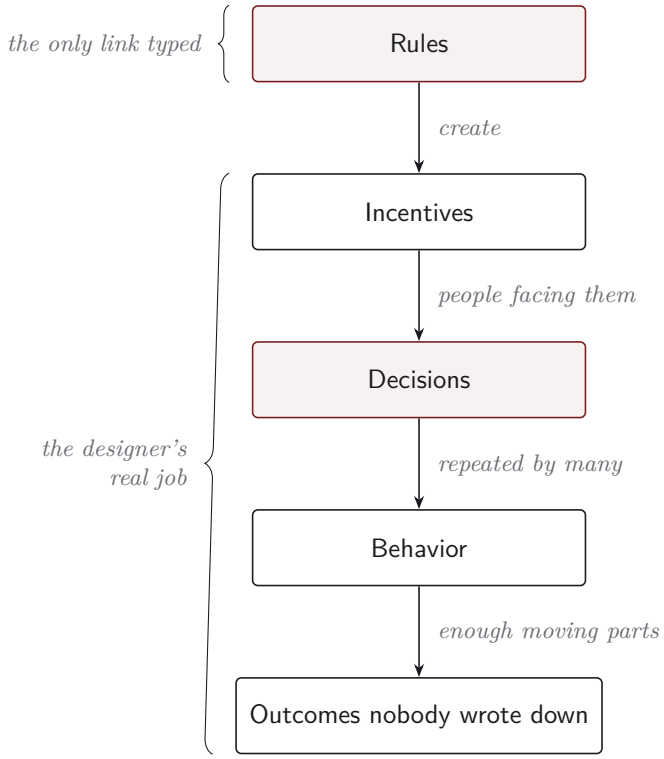


Figure 25.1. The chain under every chapter, as the introduction first stated it and the conclusion closes it. Only its first link is something a designer can type.

16. De Sousa and Sanders, interview, answer to "One of the things that we had between different bosses".
17. Staats, *The WoW Diary*, section "May 2004: The Care Bear Game".
18. Staats, *The WoW Diary*, section "October 2004: World's End".
19. Staats, *The WoW Diary*, section "Launch Day".

Video Sources

Notes that cite a video give the segment or episode label used below and a timestamp. The Candlesan segments are short videos by Wyatt Cheng on the channel `youtube.com/@candlesan`; the Distraction Makers episodes are from `youtube.com/@DistractionMakers`, hosted by Forrest Imel and Gavin Valentine. Dates are publication dates. A part A and part B under one episode number are two videos.

Candlesan

- V1. *Making an RPG with Loot? Let me save you some time.*, 2026-10-12. `https://www.youtube.com/watch?v=HEsdTuOt-Vk`
- V2. *I designed Gruul the Dragonkiller to feel like Musical Chairs.*, 2026-10-05. `https://www.youtube.com/watch?v=LsY3FHxBVgw`
- V3. *I Designed Prophet Skeram from Temple of Ahn'Qiraj*, 2026-09-07. `https://www.youtube.com/watch?v=o0dR_AaEstw`
- V4. *How to control randomness in your game*, 2026-08-10. `https://www.youtube.com/watch?v=2anY9ICA-Pg`
- V5. *What's a "Set Monster"? (Idea for Diablo used in WoW instead)*, 2026-08-06. `https://www.youtube.com/watch?v=gcwqvVa_UQM`
- V6. *Designing Loatheb in Naxxramas*, 2026-07-30. `https://www.youtube.com/watch?v=NksxgFq6eAg`
- V7. *The Simple Math Hiding in Your Favorite Games*, 2026-09-21. `https://www.youtube.com/watch?v=IzHAjV8YEW4`

Distraction Makers

- Episode 1. *The Mechanic That Changed Magic Forever*, 2024-03-26. `https://www.youtube.com/watch?v=u4TGSdkKqxc`

- Episode 2. *The Mechanic That Violated the Laws of Magic*, 2024-04-02. `https://www.youtube.com/watch?v=7xD_yfwRtE4`
- Episode 3. *The Foundational Mechanics of Magic: The Gathering*, 2024-03-19. `https://www.youtube.com/watch?v=Y6nZR7ZBD_w`
- Episode 4. *Why Does Magic: The Gathering Have Two Main Phases?*, 2024-02-22. `https://www.youtube.com/watch?v=IfL3RO1o58U`
- Episode 5. The marketing-experience episode, not among the channel's published videos.
- Episode 6. *Competitive Games are Struggling – What Happened?*, 2026-04-23. `https://www.youtube.com/watch?v=8iA8eYQARio`
- Episode 7. *How Boosters Solved One Of Magic's Biggest Problems and Why They Don't Anymore*, 2024-05-16. `https://www.youtube.com/watch?v=sGx6ZRmSeF8`
- Episode 8. *What is Magic Even Doing?*, 2026-01-08. `https://www.youtube.com/watch?v=K7pcVRF_J3Q`
- Episode 9. *Why Reading The Card Doesn't Explain The Card Anymore*, 2024-05-02. `https://www.youtube.com/watch?v=CJFUVI99yQg`
- Episode 10. *Does Magic Need Standard?*, 2025-09-09. `https://www.youtube.com/watch?v=LSjBzKO3FEQ`
- Episode 11. *Magic is Falling Apart – Is Universes Beyond to Blame?*, 2025-10-09. `https://www.youtube.com/watch?v=YVDv2QFI3nc`
- Episode 12. *The Overlooked Influence of Mana In Magic:The Gathering*, 2024-07-04. `https://www.youtube.com/watch?v=b9Q_zqtXPU4`
- Episode 13, part A. *Why Luck is so Important in Magic: The Gathering*, 2024-08-08. `https://www.youtube.com/watch?v=CKZ38X2No1s`
- Episode 13, part B. *How Players Break Games and What Designers Can Do To Stop Them*, 2025-04-08. `https://www.youtube.com/watch?v=NKXUsXI3zkU`
- Episode 14. *Why do Game Designers Keep Blaming Players for Their Failures?*, 2026-03-19. `https://www.youtube.com/watch?v=Y2hNxEzY6i8`

- Episode 15, part A. *Why Do Magic Cards Have So Much Text Now?*, 2024-03-12. https://www.youtube.com/watch?v=QMmMYy33lUA
- Episode 15, part B. *Why Is The Strategy Of Magic: the Gathering So Different?*, 2024-02-13. https://www.youtube.com/watch?v=yVRksML4Vas
- Episode 16. *Brackets Can't Fix This – How Magic's Systems Test Player Limits*, 2026-02-05. https://www.youtube.com/watch?v=PuS83zsKHo4
- Episode 17, part A. *Old Magic is Still Awesome | Premodern*, 2026-02-24. https://www.youtube.com/watch?v=ZQZ0FXo75Bs
- Episode 17, part B. *The Biggest Lesson Magic: The Gathering Taught Game Designers*, 2023-11-30. https://www.youtube.com/watch?v=SV_SFjCAxUE
- Episode 18. *Is This Magic's Biggest Design Mistake?*, 2025-10-16. https://www.youtube.com/watch?v=LGuor76-JI0
- Episode 19. *Why Magic: The Gathering Is The Most Complex Game Ever Made*, 2024-06-04. https://www.youtube.com/watch?v=dxTXQRIc-0s
- Episode 20. *Decoding The Unspoken Rules of Games*, 2025-01-02. https://www.youtube.com/watch?v=SMy69uQ8mIk
- Episode 21, part A. *The Biggest Mistake of Strategy Games*, 2024-11-26. https://www.youtube.com/watch?v=XM63oUZieEQ
- Episode 21, part B. *We Were Wrong About The Paradox of Skill*, 2025-04-15. https://www.youtube.com/watch?v=1SHKaxyTVJk
- Episode 22, part A. *Were Counterspells a Mistake?*, 2025-05-27. https://www.youtube.com/watch?v=yBe3taQLlpE
- Episode 22, part B. *How Game Designers Create Bluffing*, 2025-06-19. https://www.youtube.com/watch?v=E_9thxABVKM
- Episode 23. *It's Time to Talk About Proxies*, 2026-08-25. https://www.youtube.com/watch?v=MvxzvzsmVG8