Independent Research

A Mathematical Blueprint for Sustainable Success on Roblox

Player Dynamics, Causal Measurement, Constrained Growth, and Self-Renewing Live Operations

Nishanth Dasari

Roblox Developer, Game Owner and Operator
Template Seller and Contributor to Top Roblox Experiences
Former Teen Council Member
Age 16
Not affiliated with Roblox Corporation

April 14, 2026
Published: October 26, 2026

Figures, Sources, and Reproducibility

Owner-operator perspective and interests. Nishanth Dasari owns and operates Push Party. Section 20 uses its first-party Creator Dashboard and Ads Manager aggregates. The author has a commercial interest in that game. Murder Mystery 2 and Natural Disaster Survival are external comparisons; no ownership of those games or access to their private telemetry is claimed.

Figures and calculations. The fifteen figures are vector graphics compiled from PGFPLOTS and TikZ/PGF source [32, 31]. Curves use the stated formulas, parameters, or reported audit values. Analysis scripts use NumPy and SciPy [33, 34]. Synthetic illustrations are not observed player outcomes. Section 30.4 uses seed 20261027; its reported losses are expected synthetic regret, with Monte Carlo standard errors.

Data coverage and access. The author's operating perspective draws on personal statistics since 2022, Push Party statistics beginning October 21, 2026, and historical observations of games dating to the early 2010s. The quantitative audit in Section 20 is a specific October 2026 snapshot. The broader history is background, not a supplied longitudinal dataset. October 26 is the original documentation-check and audit date.

Public materials. The typeset manuscript: download the PDF; the corresponding LaTeX source and calculation scripts are at download source and scripts. These materials reproduce the mathematical and synthetic examples. The underlying private dashboard exports are not included, so the audit cannot be independently reconstructed from this package. Aggregate observations are not causal estimates.

Study and publication status. Section 31 is a proposed prospective protocol. No completed preregistration, recruitment, experiment, or institutional ethics approval is asserted. Before implementation, applicable ethics, privacy, consent, and platform requirements must be resolved. This manuscript presents an independent theoretical synthesis and a provisional descriptive audit; its crystal model has not been empirically validated.

Abstract

Live-game operators must connect acquisition, repeat play, technical reliability, and limited development capacity when choosing what to change after launch. This monograph develops an operational framework that links these decisions through explicit outcome definitions, cohort accounting, survival analysis, causal measurement, and constrained optimization. A descriptive first-party Push Party launch audit supplies observed acquisition, retention, advertising, and error statistics, illustrating dashboard disagreement and incomplete attribution. A proposed 28-facet crystal representation connects declared design scores to retained click-through. Conditional derivations apply classical water-filling allocation and graph-Laplacian regularization to that representation, while a synthetic creative-selection experiment compares click optimization with delayed retained-conversion learning. In the stated simulated environment, the delayed-retention policy reduces expected retained-player loss by approximately 43 percent relative to uniform allocation. A targeted related-work analysis positions these components against existing game-survival, churn-prediction, game experimentation, graph learning, and delayed-feedback research. The contribution is a decision-oriented synthesis, an observed operating case, and explicit modeling choices with reproducible conditional examples. A prospective protocol sets out a possible empirical test of one facet-repair prediction. Predictive and causal validation of the full crystal representation remains future work.

Keywords: Roblox; game development; multi-armed bandits; pre-registration; retention; survival analysis; causal inference; stochastic control; live operations; game economies.

Contents

Related work and contribution boundaries iv

Notation and model domains vii

Who this is for and how to read it 1

1 Success as a controlled trajectory 3

2 Evidence, denominators, and measurement contracts 5

3 Cohort renewal and the population accounting equation 7

4 Multistate dynamics: departure is not permanent churn 8

5 First-session survival and the calculus of friction 10

6 Onboarding as information acquisition, not instruction volume 11

7 Challenge, progression, and the geometry of meaningful choices 13

8 Technical reliability as an endogenous retention constraint 15

9 Social play, network effects, and interference 17

10 Economy conservation, inflation, and progression integrity 19

11 Monetization as retained value under uncertainty 21

12 Advertising, creative selection, and marginal acquisition 23

13 Discovery as an uncertain distribution response 25

14 Causal experiments for games with changing audiences 26

15 Statistical inference, shrinkage, and sequential decisions 28

16 Maturity, missingness, and delayed outcome accounting 30

17 Production capacity, technical debt, and the marginal feature 32

18 A unified constrained control formulation 33

19 Synthetic experiments and falsifiable predictions 35

20 A preliminary launch audit: Push Party 37

21 Analytical limits and useful propositions 40

22 The event system and an end-to-end experimental protocol 45

23 An illustrative new-game design: the orchard escape loop **47**

24 Release engineering and the bridge from a patch to an experiment **48**

25 Novelty, memory, and the production of new situations **50**

26 Update cadence and stability of the retention system **51**

27 Murder Mystery 2 as a structural retention case **54**

28 Designing updates that continue working after release **56**

29 The crystal model: a 28-facet synthesis **59**

30 Beyond one game: extensions and open problems **66**

31 A proposed prospective test on Push Party **71**

32 Conclusion **73**

References **75**

Related Work and Contribution Boundaries

Research-context revision: October 29, 2026. This targeted review positions the manuscript's central measurement and decision models. It is not an exhaustive systematic review, and absence from the papers discussed here does not establish priority. The research question is: how can a live-game operator connect acquisition, retained play, and constrained development choices without confusing descriptive observations, conditional model implications, and causal effects?

Game retention and survival measurement

Viljanen et al. [39] apply survival analysis to game playtime, including censoring-aware summaries and an A/B comparison, with a mobile-game example. This is direct prior work for Sections 4–5 and 14–16. Survival curves, censoring, and uncertainty-aware playtime comparisons are therefore methodological foundations of this manuscript, not contributions originating here. The present framework connects those tools to cohort accounting and operational resource allocation, but does not demonstrate a more accurate survival estimator.

Rothmeier et al. [40] compare four labeling approaches and eight classification methods for churn and disengagement in The Settlers Online, and evaluate on an independent dataset. That comparison shows why the target definition and validation design matter. Sections 2–4 adopt explicit outcome contracts and distinguish absence, departure, and return. The Push Party audit reports aggregate observations; it does not train or validate a player-level churn predictor, and its retention percentages are not directly comparable to a classifier's AUC or accuracy.

Game experiments and social interference

Viljanen et al. [41] use recurrent-event Cox models to compare retention and purchase rates in heterogeneous game cohorts with different follow-up times. The manuscript's use of mature cohorts and declared estimands builds on this existing line of game-experiment research. A Cox analysis, a session-survival curve, and exact-day D1 retention answer different questions; no ranking among them follows merely from their numerical values.

Ugander et al. [42] analyze graph cluster randomization and exposure-weighted estimation under social interference. Section 9's concern about mixed social environments and Section 14's discussion of clustered assignment belong to this established experimental-design problem. A server is not automatically an independent cluster: parties, repeat visits, and cross-server ties can connect assigned units. The proposed study must define its exposure model and assignment unit before enrollment; observing server arm composition alone does not eliminate interference bias.

Facet repair and classical allocation

The objective in Theorem 29.4 is a bounded, weighted separable logarithmic allocation problem. Classical water-filling is developed, for example, by Boyd and Vandenberghe [43], Example 5.2 in Section 5.5.3. The theorem here applies the same KKT mechanism to declared facet scores, unequal costs, score ceilings, and zero objective weights. A self-contained proof makes the assumptions inspectable; it does not establish invention of water-filling or a new general-purpose allocation algorithm. The model-specific choice is what the scores, costs, and objective represent.

The numerical contrast between clarity-optimal and retained-click-through-optimal repair is a consequence of the stated synthetic profile. Its practical value is the explicit demonstration that changing an objective changes an allocation. Neither the 36 percent modeled increase nor the 26 percent comparison is an observed treatment effect. Real repair effort can have uncertain, nonlinear, and interacting effects on scores; that response mapping must be estimated before the formula can justify actual production spending.

Graph regularization and the role of geometry

Belkin, Niyogi, and Sindhwani [44] develop geometry-based regularization within a learning framework. Their graph structure concerns examples; the crystal penalty in Section 29.7 concerns coefficients indexed by facets. These are different statistical objects, but both use graph smoothness to constrain a fit. The present choice of a prism graph is a declared prior, not a graph learned from game outcomes. Uniqueness of the penalized solution does not imply that the true 28 effects are identified or that the chosen adjacency predicts well.

An empirical comparison should retain a held-out set of games or versions, fit score transformations and tuning parameters only on training data, and compare the prism penalty against ridge, a group-level model, and prespecified alternative or permuted adjacencies. If versions share players or common development history, split at an appropriately independent unit rather than treating correlated versions as independent replications. A training fit or favorable spectrum alone cannot establish that the geometry earns its statistical role.

Delayed outcomes and creative-selection policies

Joulani, Gyorgy, and Szepesvari [45] study online learning with delayed feedback and give delay-aware methods and regret analysis. Delayed feedback is therefore an existing learning problem, not a new algorithmic contribution of Section 30.4. That section is a synthetic illustration of reward misalignment: an invented retention profile makes the click winner differ from the retained-conversion winner. The simulation does not compare against all existing delay-aware policies, nor does it estimate a causal advertising lift in Push Party.

An informative extension would vary retention-rate ordering, delay, conversion sparsity, horizon, and proxy calibration; report Monte Carlo uncertainty; and use an exploration-constrained oracle as well as an unconstrained oracle. It should compare policies optimizing the same target before attributing improvement to the learning algorithm. Equal impression allocation is not equal spending when creative prices differ. Those extensions are evaluation requirements, not completed results claimed in this revision.

What this manuscript contributes

The manuscript offers three connected pieces of work, with different evidence requirements:

1. **An operational synthesis.** Sections 1–28 connect established mathematical tools through explicit outcomes, units, and resource constraints for live-game decisions. The value claimed is the integration and declared decision contracts; the individual tools are attributed to prior work.

2. **A descriptive first-party case.** Section 20 documents observed Push Party acquisition, engagement, advertising, and operational statistics. It supports a specific account of

dashboard disagreement and incomplete attribution. A single observational audit does not establish transportability, an intervention effect, or validation of the crystal.

3. **A proposed facet representation with conditional applications.** Sections 29–31 specify a 28-facet model, derive consequences using established methods, illustrate choices synthetically, and propose a prospective test. The taxonomy, score contracts, and retained-conversion mapping are modeling choices to evaluate. This review does not establish that no similar model exists or that the representation improves prediction or decisions.

Evidence classes and the claims they can support.

Material	Supported claim	Claim not established
Push Party audit	Reported aggregate observations in one launch window	Causal lift, cross-game validity, or complete crystal validation
Conditional derivations	Mathematical implications under stated assumptions	Novelty of the underlying methods or validity of player-behavior assumptions
Synthetic experiments	Outcomes in the declared simulated environment	Observed retention improvement or general policy superiority
Prospective protocol	A specified route to a future comparison	Completed preregistration, recruitment, or experimental results

The strongest currently supported position is thus a research monograph combining methodological synthesis, an observed operating case, and an unvalidated model proposal. A claim of a new predictive or causal method would require an explicit improvement over relevant alternatives and suitable evidence. Adding citations clarifies that boundary; it cannot supply missing experiments.

Notation and Model Domains

Symbols are local to a model when explicitly redefined. In particular, a daily participation probability is not an instantaneous occupancy, and a session-survival probability is not exact-day retention. This distinction is necessary for interpreting the equations consistently.

Table 1. Principal notation.

Symbol	Meaning	Units or domain
$t,\ c,\ a$	Calendar time, arrival cohort, cohort age	Days unless specified
$D_t,\ M_t$	Distinct daily and trailing-month active users	Players
A_c	Newly acquired distinct users in cohort c	Players
$r_c(a)$	Exact-day participation probability	Dimensionless
L_H	Expected player-days through age H	Player-days per acquisition
$T,\ S(t),\ h(t)$	First-session duration, survival, hazard	Minutes; probability; inverse minutes
$P,\ Q$	Daily transition matrix, continuous-time generator	Stochastic; inverse time
$n(a,t,z)$	Expected cohort-state population in segment z	Players per cohort-age unit
$b,\ B$	Advertising spend, available spending budget	One declared currency
$g,\ m,\ \mu$	Receipts or conditional value, as specified	Currency per period or active user
$q_t,\ u_t$	Product state, feasible control action	Model-specific vectors
$h_j,\ H_t$	Production allocation and available hours	Hours
$\delta,\ \rho$	Discount factor or rate; locally redefined	Dimensionless or inverse time
W_i	In-game wallet balance	Declared game currency
N_i	Remaining meaningful novelty proxy	Model-specific value units
τ	Update interval; sometimes a treatment contrast	Days or stated outcome units
$\theta,\ \mathcal{D}$	Unknown parameters, observed evidence	Model-specific
$\varrho(B)$	Spectral radius of a population-cycle operator	Dimensionless
$\mathcal{F}_t$	Information available at decision time	Sigma-algebra
$\mathbb{E},\ \mathrm{Var},\ \mathrm{Cov}$	Expectation, variance, covariance	Induced by defined random variables

Figures. Every figure is a vector drawing produced from stated formulas or reported data with PGFPLOTS and TikZ; the software and numerical methods are listed at the end of the References.

Model boundaries. Equations for first-session survival concern uninterrupted sessions. Renewal equations concern daily participation. Markov and transport equations are proposed population approximations. Queueing formulas are illustrative stress models, not reverse-engineered Roblox internals. Optimization results are conditional on estimated response functions and declared constraints.

Who This Is For and How to Read It

This manuscript is written for the people who decide what happens to a Roblox experience after it ships. If you build, run, fund, or study live games, one of the reading paths below is meant for you. You do not need to read all thirty-two sections in order, and you do not need every derivation to use the results.

Four readers. The work has four intended audiences, and each gets something different from it.

1. **Developers and small teams.** You make the game and usually also run its ads, its updates, and its analytics. You need to know which problem to fix first when players leave, whether a high click-through rate is good news, and when to stop spending. Start with the Practical implication at the end of each section; each one states the decision the mathematics supports in a few sentences.
2. **Studio leads and producers.** You allocate scarce hours and budget across features, campaigns, and sometimes several games. You need decision rules that survive uncertainty, an honest reading of launch dashboards, and a way to tell a lasting improvement from a one-time spike.
3. **Analysts and data-minded developers.** You own the numbers. You need precise definitions, comparisons that hold when the audience changes, and protection against reading noise as signal in small cohorts.
4. **Researchers and students.** You may know survival analysis, causal inference, or control theory but not live games. You need the models stated precisely, the assumptions made explicit, and open problems you can work on. Sections 30 and 31 are written mainly for you.

Table 2. Suggested reading paths. Every path ends with Section 29, the crystal model, which ties the structures together.

Reader	Start with	What you leave with
Developer or small team	1, 5, 6, 8, 20, 23, 28, 29, 31	What to fix first, how to read a launch honestly, and how to design updates that keep working
Studio lead or producer	1, 3, 4, 12, 17, 18, 26, 29, 30	Budget and production allocation rules, cadence, and when a game has earned more investment
Analyst	2, 14, 15, 16, 19, 20, 22, 29, 31	Measurement contracts, experiment design, small-sample inference, and an event system that can answer the questions
Researcher or student	3–5, 18, 21, 26, 29–31	Formal models, propositions with proofs, limits of identification, and open problems

What you need to know. Each section opens in plain language before it introduces notation, and closes with a Practical implication that needs no mathematics. Following the derivations requires undergraduate probability and some calculus; Sections 4, 18, 21 and 26 also use linear algebra. Every symbol is defined where it first appears and collected in the notation table above.

What this is not. It is not a list of tricks for the algorithm, a promise of revenue, or a reverse-engineering of Roblox systems. It is a set of models for reasoning about a game you control, with the evidence each claim needs and the observations that would prove it wrong.

1 Success as a controlled trajectory

A game succeeds over time, rather than at a single instant. A large launch can conceal weak return behavior. A quiet week can conceal a small, durable community. A profitable event can accelerate future churn if it damages fairness, trust, or the economy. Consequently, success requires a time horizon, a population, a resource constraint, and a definition of player value. A dashboard number without these ingredients is a description, not a decision criterion.

Consider a studio choosing actions $u_0, u_1, \ldots, u_T$. An action may allocate engineering effort, change a first-session interaction, release a map, adjust a progression curve, select a creative, or reserve cash rather than spend it. Let X_t summarize the product and community state. The studio observes a noisy projection Y_t, not the entire state:

$$X_{t+1} = F_\theta(X_t, u_t, \xi_{t+1}), \tag{1.1}$$

$$Y_t = H_\theta(X_t, \varepsilon_t). \tag{1.2}$$

Unknown θ includes preferences, transition rates, technical sensitivities, and distribution responses. Disturbances ξ_t include competing releases, holidays, outages, and the arrival of particular social groups. Measurement error ε_t includes delayed attribution, incomplete sessions, aggregation, and instrumentation failure.

The object of optimization is a policy π: a rule choosing actions from information available at the time. A policy cannot legitimately depend on a cohort's future revenue before that revenue exists. If $\mathcal{F}_t$ is current information, u_t must be $\mathcal{F}_t$-measurable. This prevents hindsight evaluation in which the eventual winning creative is treated as though it had been known before testing.

1.1 Outcomes before a scalar objective

Let outcomes include retained player-days P, meaningful social interactions S, net receipts G, player-reported experience V, and operating burden C. The feasible set is

$$\mathcal{A} = \{(P, S, G, V, C) : \text{attainable under declared constraints}\}. \tag{1.3}$$

A Pareto-efficient design admits no improvement in one desired coordinate without deterioration elsewhere. There need not be one design maximizing every coordinate. A studio should identify tradeoffs before hiding them behind numerical weights.

A scalar planning objective is nevertheless useful:

$$J(\pi) = \mathbb{E}_\pi\left[\sum_{t=0}^{T} \delta^t \{g_t - b_t - c_t + \omega v_t\}\right] - \lambda_{\text{risk}}\mathcal{R}(\pi), \qquad 0 < \delta \leq 1. \tag{1.4}$$

Here g_t, b_t, c_t use the same monetary unit; v_t is a declared player-value measure; and ω converts its units to objective units. Without a defensible ω, report a Pareto frontier. Robux, USD, and player-days must not be added without conversion conventions.

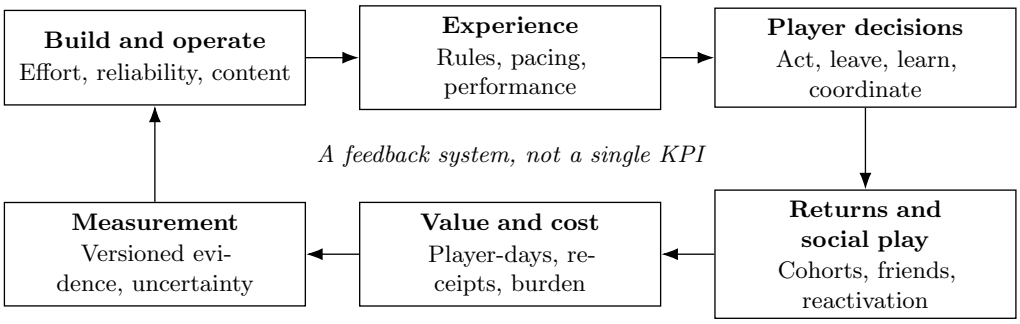


Figure 1. The proposed feedback architecture. Each box is a model component and each arrow a modeled dependency to investigate; the arrows form a single cycle, which is the sense in which no one indicator can be read in isolation. Arrows are not estimates of causal effects. Drawn by the author with TikZ [31].

1.2 Feasibility is part of the product

Reliability, accessibility, and capacity enter as constraints:

$$\Pr(\text{critical save failure} \mid z) \le \epsilon_z, \tag{1.5}$$

$$\mathbb{E}[\text{engineering hours in week } t] \le H_t, \tag{1.6}$$

$$\sum_t b_t \le B, \qquad b_t \ge 0. \tag{1.7}$$

The tolerances ϵ_z are studio requirements, not Roblox-wide numerical thresholds. Segment z may represent device class, language, or acquisition source. A design working for the average user may remain unacceptable when a large mobile segment experiences persistent failure.

Definition 1.1 (Sustainable success)**.** For a declared horizon and resource envelope, sustainable success means maintaining a feasible policy that creates positive player value and a viable operating trajectory, with uncertainty reported and without depending on indefinitely increasing paid acquisition.

A niche game can satisfy this definition. It does not require the largest audience or an institutional budget. It also permits a rational stop: if the next dollar or production week has lower expected value here than elsewhere, conserving resources can be the best continuation policy.

1.3 A blueprint is an identification problem

Writing F_θ does not identify θ. Many parameter settings can produce the same DAU curve. An advertising surge followed by weak returns may resemble stable acquisition followed by an outage. Observational equivalence means that model complexity must be earned with additional measurements or interventions.

The practical object is therefore a sequence

$$\text{hypothesis} \longrightarrow \text{measurable contrast} \longrightarrow \text{intervention} \longrightarrow \text{updated belief}. \tag{1.8}$$

A design decision is justified by how it changes expected future outcomes under uncertainty. A model earns its place by improving that decision, not by displaying sophisticated notation.

Practical implication. Start with a ledger specifying what will change, which outcome should move, whom it should help, how long evidence takes to mature, and what observation would reverse the decision. The rest of the manuscript develops that ledger mathematically.

2 Evidence, denominators, and measurement contracts

A dashboard can establish that a percentage changed. It cannot alone establish why. Source mix, version exposure, weekends, device distribution, and missing telemetry can change the same metric. The first analytical task is to distinguish a measurement from its explanation.

This manuscript separates four claim classes. An *Observation* is directly inspected evidence. A *Model* is a proposed structure. A *Hypothesis* is a testable explanation. A *mathematical result* follows logically from assumptions. A proof about a Markov model does not prove that a game obeys that model.

2.1 Cohorts and exact-day retention

For cohort c with N_c distinct new players, let $I_{ic}(d)$ indicate that player i plays on cohort-relative day d. Exact-day retention is

$$r_c(d) = \frac{1}{N_c} \sum_{i=1}^{N_c} I_{ic}(d). \tag{2.1}$$

The same initial cohort supplies the denominator at every horizon. Returning on day 7 after missing day 6 contributes to $r_c(7)$ but not $r_c(6)$. Exact-day retention therefore need not decrease with d. Its graph is not a continuous survival curve.

For distinct daily active users D_t and distinct users M_t active in a specified trailing monthly window, stickiness is

$$\sigma_t = \frac{D_t}{M_t}. \tag{2.2}$$

The monthly denominator is a set union, not the sum of DAU. Near launch, most monthly users may have arrived today. High σ_t can coexist with low repeat play; it is not a cohort return probability.

2.2 Averages with different weights

With session durations L_j, distinguish

$$\bar{L}_{\text{session}} = \frac{\sum_j L_j}{\#\text{sessions}}, \tag{2.3}$$

$$\bar{L}_{\text{player}} = \frac{\sum_j L_j}{\#\text{distinct active players}}. \tag{2.4}$$

A daily-average card may average the daily ratios, producing a third estimator. Roblox's engagement and retention documentation should anchor dashboard interpretation [12, 11]. Custom pooled analysis must retain the original definition.

Proposition 2.1 (Average-of-ratios discrepancy)**.** *For totals a_t and positive denominators n_t,*

$$\frac{1}{T}\sum_t \frac{a_t}{n_t} - \frac{\sum_t a_t}{\sum_t n_t} = -\frac{\mathrm{Cov}_T(n_t, a_t/n_t)}{\bar{n}}, \tag{2.5}$$

where Cov_T *uses divisor* T *and* $\bar{n} = T^{-1}\sum_t n_t$.

Proof. Put $q_t = a_t/n_t$. The identity $\overline{nq} = \bar{n}\bar{q} + \mathrm{Cov}_T(n, q)$ implies the result after division by $\bar{n}$. □

A three-player prelaunch day with long playtime can disproportionately influence an unweighted card. A high-volume launch day with brief sessions should not receive the same inferential weight simply because both occupy one calendar date.

2.3 The measurement contract

Every measure should retain

$$\mathcal{M} = (\text{entity}, \text{event}, \text{denominator}, \text{window}, \text{timezone}, \text{source}, \text{version}, \text{maturity}). \tag{2.6}$$

A report saying D1 is 8% is incomplete. A report specifying exact-day D1 for newly acquired sponsored players first joining a named build in a UTC interval, after enough follow-up, defines an estimand.

Retain event time and ingestion time separately. If $\ell_e = t_e^{\text{ingest}} - t_e^{\text{event}}$, a snapshot at T misses events with $t_e^{\text{event}} \le T < t_e^{\text{ingest}}$. Faster ingestion can imitate improved activity in recent windows. Missing events should be distinguished from explicit negative outcomes.

2.4 Attribution is not incrementality

An attributed purchase is credited to an advertisement under reporting rules. An incremental purchase would not have happened without the intervention. With potential outcomes $Y_i(1), Y_i(0)$, the latter estimand is

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)]. \tag{2.7}$$

Attribution observes neither pair for one player. It is useful accounting, not automatic causal identification. This matters especially when ads reach existing players who may have returned anyway.

2.5 Instrument success, failure, and exposure

A purchase prompt, a successful receipt, and a delivered entitlement are different events. A shop-open count cannot stand in for delivered purchases. Similarly, a server's configured feature flag is not proof that every client saw the feature. Log assignment, actual exposure, success, and failure as distinct stages.

A measurement audit should ask whether totals reconcile across independently generated records. Perfect equality is not always expected because units differ; explain expected differences before treating unexplained discrepancies as growth.

Practical implication. When two dashboards disagree, compare contracts before choosing the more flattering number. Correct denominator selection often resolves more uncertainty than another layer of modeling.

3 Cohort renewal and the population accounting equation

Acquisition and retention jointly generate daily activity. Let A_c be players first arriving on day c, and $r_c(a)$ their probability of playing at cohort age a. Then

$$\mathbb{E}[D_t] = D_t^{\text{old}} + \sum_{c=0}^{t} A_c \, r_c(t-c). \tag{3.1}$$

Old users first arriving before the measured window contribute D_t^{old}. Independence is unnecessary for this expectation: linearity suffices. Dependence becomes important when calculating uncertainty.

For stationary cohorts $r_c(a) = r(a)$, acquisition is convolved with a return kernel. Large advertising pulses create large activity pulses even when the kernel is weak. A DAU plot cannot separately identify input quality and return behavior.

3.1 Player-days per acquisition

Expected player-days over horizon H are

$$L_H = \sum_{a=0}^{H} r(a). \tag{3.2}$$

When $r(0) = 1$, $L_H \geq 1$. This is not continuous lifetime in the game; it is the expected number of active days. Constant acquisition A eventually produces, under stationarity and summability,

$$\mathbb{E}[D] \approx AL_\infty, \qquad L_\infty = \sum_{a \geq 0} r(a) < \infty. \tag{3.3}$$

If two synthetic games acquire 1,000 users daily but have L_{30} equal to 1.4 and 3.1, their within-horizon activity contributions approach 1,400 and 3,100 DAU. Doubling advertising for the first does not improve its return kernel.

3.2 A sensitivity decomposition

When a quality parameter q changes both acquisition and repeat play,

$$\frac{\partial \mathbb{E}[D_t]}{\partial q} = \sum_{c \leq t} A_c \frac{\partial r_c(t-c)}{\partial q} + \sum_{c \leq t} \frac{\partial A_c}{\partial q} r_c(t-c). \tag{3.4}$$

The first term concerns players already acquired; the second concerns changed reach. A before-and-after DAU increase mixes both. Randomized onboarding treatment with shared acquisition conditions isolates a more defensible portion of the first effect.

The convolution also has capacity limits. More players can improve social matching, then worsen latency or overcrowd encounters. Write

$$r_c(a) = r(a; X_{c:c+a}, z_c) \tag{3.5}$$

when the experienced population path influences return behavior. A linear extrapolation from a tiny test then becomes a local approximation, not a scaling law.

3.3 DAU and concurrency

Let ν be sessions per active player per day and ℓ pooled average session minutes. Average concurrency follows the player-minute identity

$$\bar{C} \approx \frac{D\nu\ell}{1440}. \tag{3.6}$$

It estimates time-average CCU, not peak CCU. Compatible session boundaries and weights are required. Little's law supplies the corresponding arrival-rate-times-duration relation [4]. A synthetic game with 10,000 DAU, 1.1 sessions per user, and three-minute sessions averages roughly 23 concurrent players. Large reach can therefore coexist with low visible population.

Peak concurrency additionally depends on time-of-day synchronization. A global audience spread across time zones produces a different peak-to-average ratio from a school-age audience concentrated in one region. Multiplying average CCU by an arbitrary constant to claim peak demand conceals this uncertainty.

3.4 The tail after paid acquisition stops

If paid acquisition ends at t_0, earlier cohorts remain:

$$\mathbb{E}\big[D_{t_0+h}^{\text{paid cohorts}}\big] = \sum_{c \le t_0} A_c^{\text{paid}}\, r_c(t_0 + h - c). \tag{3.7}$$

This justifies allowing results to mature after spending stops. Organic arrivals and prior-player returns remain separate labels. Otherwise the analyst may credit the entire remaining population to the final campaign.

A stopping decision therefore has two clocks: the last delivery time and the observation horizon. The first ends new exposure; the second determines which cohort outcomes are observable. Waiting for evidence does not require purchasing additional traffic.

Practical implication. Forecast an interval of plausible tails using several kernels. If optimistic and pessimistic tails imply different decisions, the next step is to reduce uncertainty about return behavior, not to extrapolate the most optimistic curve.

4 Multistate dynamics: departure is not permanent churn

Continuous survival describes an uninterrupted first session. It cannot by itself describe leaving today and returning tomorrow. Introduce states A (active), D (dormant but reachable), and C (effectively lost within the planning horizon). Lost is an operational category, not a claim that someone can never return.

For a continuous-time illustration, let a row vector $p(a)$ evolve as

$$p'(a) = p(a)Q(a), \qquad p(0) = (1, 0, 0), \tag{4.1}$$

with generator

$$Q = \begin{pmatrix} -(\alpha+\chi) & \alpha & \chi \\ \beta & -(\beta+\gamma) & \gamma \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.2}$$

All rates have units inverse time. α describes active-to-dormant transition, β reactivation, and χ, γ exit into the lost state. For constant Q, $p(a) = p(0)e^{Qa}$. For time-varying Q, the transition operator is time ordered: $e^{\int Q}$ is generally incorrect because generators need not commute.

4.1 Age and calendar time

Let $n(a, t, z)$ be an expected cohort-state vector. Aging and state transitions combine in

$$\partial_t n + \partial_a n = nQ(a, t, z; u_t), \tag{4.3}$$

$$n(0, t, z) = A_z(t)e_A^\top. \tag{4.4}$$

Instantaneous active occupancy is $\sum_z \int n_A(a, t, z)\,\mathrm{d}a$. Daily dashboard activity instead means playing at least once during a day. A model-fitting procedure must map the two definitions rather than identify them by notation.

A discrete daily model is often easier:

$$n_{a+1,t+1,z} = n_{a,t,z}P_{a,t,z}, \qquad n_{0,t,z} = A_{t,z}e_A^\top. \tag{4.5}$$

P is row stochastic. Estimating each entry independently can yield negative probabilities or row sums above one. Multinomial-logit parameterization enforces feasibility:

$$P_{ij} = \frac{\exp(\eta_{ij})}{\sum_k \exp(\eta_{ik})}. \tag{4.6}$$

4.2 Updates intervene on different transitions

An update can increase reactivation β, reduce active departure α, or worsen effective loss χ if it breaks saved progression. A notification click surge primarily informs reach and return; it does not establish improved persistence after return.

Suppose an update at t_k moves fraction ρ_k of dormant players into activity:

$$n_A(t_k^+) = n_A(t_k^-) + \rho_k n_D(t_k^-), \tag{4.7}$$

$$n_D(t_k^+) = (1 - \rho_k)n_D(t_k^-). \tag{4.8}$$

Player mass is conserved. Counting this reactivated mass as new acquisition would overstate acquisition efficiency.

Proposition 4.1 (A pulse is not a kernel improvement)**.** *If the update changes only the one-time jump ρ and leaves subsequent Q unchanged, the incremental state vector after elapsed time s is*

$$\Delta n(s) = \rho\, n_D(t_k^-)(e_A^\top - e_D^\top)e^{Qs}. \tag{4.9}$$

Proof. The initial displacement reallocates dormant mass. Both post-jump trajectories follow the same linear differential equation; their difference follows its homogeneous transition operator. □

A large initial event lift may therefore decay without improving the underlying game. The relevant test is whether post-return transitions changed, not whether launch-day activity spiked.

4.3 Heterogeneity and apparent loyalty

A latent-class mixture has

$$r(a) = \sum_{k=1}^{K} w_k r_k(a), \qquad w_k \geq 0, \qquad \sum_k w_k = 1. \tag{4.10}$$

A slowing departure rate can arise because low-interest players leave early, selecting more interested survivors. It need not imply that the experience improves with time. Source, language, device, and social-party variables should be considered before fitting elaborate hidden classes.

There is also a practical identifiability limit. If telemetry records only daily joins, several combinations of $\alpha, \beta, \gamma, \chi$ can fit the same curve. Additional events, explicit dormant definitions, or simpler models are required. Advanced mathematics cannot manufacture information absent from the observations.

Practical implication. Choose the smallest state model that changes a decision. A careful table separating new, returning, and reactivated cohorts is preferable to an unidentifiable generator with many free parameters.

5 First-session survival and the calculus of friction

Let T be uninterrupted first-session duration. Define survival $S(t) = \Pr(T > t)$, density $f(t)$ when one exists, and hazard

$$h(t) = \lim_{\Delta \downarrow 0} \frac{\Pr(t < T \le t + \Delta \mid T > t)}{\Delta} = \frac{f(t)}{S(t)}. \tag{5.1}$$

The hazard has units departures per surviving player-minute. It is not the fraction of all new users who leave in minute t. The relation

$$S(t) = \exp\!\left(- \int_0^t h(s)\, \mathrm{d}s \right) \tag{5.2}$$

follows from $S' = -hS$. A five-minute survival statistic identifies the cumulative hazard through five minutes, not the shape of h within that interval.

5.1 Why one retention point is insufficient

If $S(5) = 0.1243$, the average cumulative-hazard rate over that interval is $-\log(0.1243)/5 \approx 0.417$ per minute. A constant hazard is one compatible model. Another places most departures at loading or spawn. A third places them at a first loss. Each produces the same endpoint with different development implications.

The restricted mean session time is

$$\mathrm{RMST}(H) = \mathbb{E}[\min(T, H)] = \int_0^H S(t)\, \mathrm{d}t. \tag{5.3}$$

It avoids extrapolating an unobserved long tail and is often a more useful experimental endpoint than an unstable unbounded mean. Users still playing at the data cutoff are right censored; they are not zero-duration departures.

Kaplan–Meier estimation accommodates right censoring under an appropriate independent-censoring condition [2]. At event times t_j, with d_j departures among n_j at risk,

$$\widehat{S}(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{n_j} \right). \tag{5.4}$$

If telemetry loss is more likely on crashing devices, censoring may be informative. A mathematically correct estimator with a false censoring assumption remains biased.

5.2 Where a hazard reduction has value

Let a retained minute at time t contribute nonnegative value weight $v(t)$, and define

$$V[h] = \int_0^H v(t) S_h(t)\,\mathrm{d}t. \tag{5.5}$$

For a small perturbation $h_\epsilon = h + \epsilon\eta$, differentiation and Fubini's theorem give

$$\left.\frac{\mathrm{d}V[h_\epsilon]}{\mathrm{d}\epsilon}\right|_0 = -\int_0^H v(t) S_h(t) \int_0^t \eta(s)\,\mathrm{d}s\,\mathrm{d}t \tag{5.6}$$

$$= -\int_0^H \eta(s)\left[\int_s^H v(t) S_h(t)\,\mathrm{d}t\right]\mathrm{d}s. \tag{5.7}$$

Thus the functional derivative is

$$\frac{\delta V}{\delta h(s)} = -\int_s^H v(t) S_h(t)\,\mathrm{d}t. \tag{5.8}$$

Reducing hazard early preserves access to every downstream opportunity. This is a conditional model result: an intervention may also change player value v, later hazards, or the audience. A forced tutorial might reduce confusion while adding delay. Both effects belong in the evaluation.

5.3 Competing departure mechanisms

Write a cause-specific decomposition

$$h(t) = h_{\text{technical}}(t) + h_{\text{confusion}}(t) + h_{\text{mismatch}}(t) + h_{\text{social}}(t) + h_{\text{voluntary}}(t). \tag{5.9}$$

The decomposition is conceptual unless exit causes are observed reliably. A client error followed by departure is not proof of a technical cause; both may result from an overloaded encounter. Randomized fixes, device-stratified analysis, and reproduction provide stronger evidence.

A proportional-hazards model can relate measured covariates x to hazard:

$$h(t \mid x) = h_0(t)\exp(x^\top \beta). \tag{5.10}$$

This standard Cox form [3] assumes proportionality over time. A mobile loading problem and a late-round balance problem may violate that assumption. Time-varying effects or separate phases may be more interpretable than a single hazard ratio.

Practical implication. Instrument loading complete, first control, first meaningful action, first reward, first loss, and first successful loop. Optimize the largest reproducible failure in that sequence while checking that a longer session still contains meaningful play.

6 Onboarding as information acquisition, not instruction volume

A player joins with uncertainty about controls, goals, danger, progression, and the promise made by the thumbnail. Onboarding should resolve the uncertainty necessary for the next meaningful action. Explaining the entire game before action can create a larger barrier than the uncertainty it removes.

Let Θ_i represent a player's uncertain understanding of a mechanic. A useful interaction supplies observation O, reducing posterior uncertainty:

$$\mathrm{IG} = H(\Theta_i) - \mathbb{E}_O\big[H(\Theta_i \mid O)\big], \tag{6.1}$$

where H denotes entropy. This is a design model, not a directly observable mental quantity. In practice, successful unaided actions, help requests, and repeated control errors provide proxies.

6.1 Instruction has an opportunity cost

Suppose instruction time is x, comprehension probability is $c(x)$, and survival through instruction is $s(x)$. The probability of reaching a comprehended first action is

$$q(x) = s(x)\,c(x). \tag{6.2}$$

At an interior optimum,

$$\frac{c'(x)}{c(x)} = -\frac{s'(x)}{s(x)}. \tag{6.3}$$

The marginal proportional comprehension gain equals the marginal proportional survival loss. More instruction is not automatically better.

For a synthetic example, take $c(x) = 1 - e^{-kx}$ and $s(x) = e^{-hx}$ with $h, k > 0$. Differentiation yields

$$x^* = \frac{1}{k}\log\left(1 + \frac{k}{h}\right). \tag{6.4}$$

This result does not prescribe a universal tutorial length. It explains why the optimal amount depends on both learning speed and departure risk. A mechanic already familiar to the audience needs less explicit explanation than an unfamiliar interaction.

6.2 Action and explanation can be interleaved

A linear tutorial with stages $1, \ldots, m$ has completion probability

$$q_{\text{complete}} = \prod_{j=1}^{m} q_j. \tag{6.5}$$

Even moderately high stage success compounds into substantial loss. Conditional stage conversion q_j must be measured among those reaching stage j, not all joining players.

An alternative presents the smallest complete loop early and teaches advanced features when relevant. This changes the graph of dependencies. If stage 4 is optional for the first reward, it should not be a required predecessor in the onboarding funnel. A directed acyclic graph of prerequisites makes this explicit.

Let $G = (V, E)$ encode required dependencies, and T_v the time to complete node v. The earliest achievable first-loop completion time satisfies a critical-path recursion:

$$C_v = T_v + \max_{u:(u,v)\in E} C_u. \tag{6.6}$$

Removing a node outside the critical path may improve visual simplicity without changing time to first loop. Removing a critical mandatory step can change that time immediately.

6.3 Promise alignment

The acquisition creative creates an expectation e_i, while the opening experience supplies realized signals x_i. A mismatch index could be

$$m_i = (e_i - x_i)^\top W (e_i - x_i), \qquad W \succeq 0. \tag{6.7}$$

This is a hypothetical latent construct. It suggests a measurable experiment: hold the opening fixed, vary accurately differentiated creatives, and compare first-action and five-minute outcomes. A high-CTR creative that attracts a mismatched audience may lower downstream conversion.

The best creative may therefore be the one producing the most satisfied qualified players per dollar, rather than the most clicks per impression. Testing the thumbnail and onboarding together can reveal an interaction that separate tests miss.

6.4 A teachability experiment

Randomize eligible new users between a brief action-first opening and the existing opening. Keep the promised mechanic, matchmaking conditions, and reward values consistent. Predeclare first-loop completion within a time limit, restricted mean first-session time, D1, and critical-error guardrails.

Record skip behavior as an outcome, not as random assignment. Players who voluntarily skip differ from those who do not. Comparing those groups directly does not identify the causal effect of skipping. Assignment to a skip-enabled design supports an intention-to-treat contrast.

Practical implication. Give players a successful use of the core mechanic before demanding memorization. An onboarding improvement is established by improved action and return outcomes, not by the amount of text removed.

7 Challenge, progression, and the geometry of meaningful choices

A core loop creates decisions whose outcomes a player can understand and influence. Progression changes future possibilities. When progression only scales numbers without changing decisions, content exhaustion can occur long before the final upgrade is purchased.

Let player skill be s and challenge difficulty d. A simple synthetic success model is

$$p_{\text{win}}(s, d) = \text{sigmoid}\left(\frac{s - d}{\tau}\right) = \frac{1}{1 + \exp\big(-(s - d)/\tau\big)}. \tag{7.1}$$

$\tau > 0$ represents outcome noise or softness of the skill boundary. This is not a measured law of Roblox behavior. It supplies a tunable language for discussing difficulty and player segmentation.

7.1 Challenge utility is not win probability alone

An illustrative encounter value is

$$U(s, d) = R\, p_{\text{win}}(s, d) - c_{\text{time}}(d) + \kappa\, p_{\text{win}}(1 - p_{\text{win}}). \tag{7.2}$$

The last term rewards uncertain but achievable outcomes. Its coefficient is a preference assumption, not a universal psychological constant. Some players prefer mastery; others prefer predictable collection. A single target win rate across all modes ignores these differences.

Learning can be modeled locally by

$$s_{n+1} = s_n + \eta\,\phi(s_n, d_n) + \epsilon_n, \tag{7.3}$$

where ϕ measures useful practice. If trivial challenges produce little learning and impossible challenges produce repeated confusion, an adaptive difficulty policy should seek informative practice rather than merely maximize immediate wins.

7.2 Upgrade pacing

Suppose upgrade k costs C_k currency and current earning rate is y_k. Ignoring stochasticity, its time cost is

$$T_k = \frac{C_k}{y_k}. \tag{7.4}$$

For geometric curves $C_k = C_0 a^k$, $y_k = y_0 b^k$,

$$T_k = \frac{C_0}{y_0}\left(\frac{a}{b}\right)^k. \tag{7.5}$$

The ratio a/b controls pacing. When $a > b$, waits grow exponentially. When $a < b$, upgrades accelerate, potentially exhausting progression. A visually modest change in either multiplier can substantially alter late-game pacing.

In continuous progression level x, let $\dot{x} = y(x)/C(x)$. The time to move from x_0 to x_1 is

$$T(x_0, x_1) = \int_{x_0}^{x_1} \frac{C(x)}{y(x)}\,\mathrm{d}x. \tag{7.6}$$

Designing a target pacing function $T'(x)$ implies $C(x)/y(x) = T'(x)$. This reverses the usual workflow: specify the intended rhythm first, then choose costs and yields consistent with it.

7.3 Risk and reward in a chase-and-collect loop

Consider a player choosing whether to take one more resource before returning to safety. Let carried value be v, extra reward r, survival probability for the extra action and escape be p, and loss cost be ℓ. Compared with banking v, the extra action is worthwhile under a risk-neutral model when

$$p(v + r) - (1 - p)\ell > v, \tag{7.7}$$

or

$$r > \frac{1 - p}{p}(v + \ell). \tag{7.8}$$

A meaningful decision requires both sides to be plausible for some player states. If the inequality is always satisfied, there is no real banking decision. If never satisfied, the extra action is decorative.

For risk-averse players with utility u, compare

$$p\,u(v + r) + (1 - p)\,u(-\ell) \quad \text{with} \quad u(v). \tag{7.9}$$

The design should provide understandable signals about risk and an opportunity to bank. It should not require hidden loss probabilities to create tension.

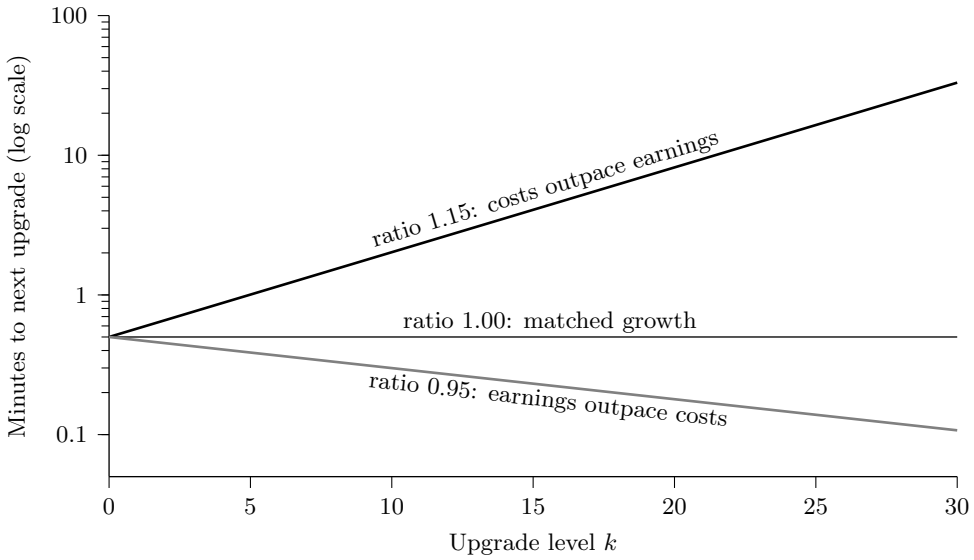


Figure 2. Synthetic pacing under geometric cost and earning curves, with initial upgrade time 0.5 minutes. The ratio of cost growth to earning growth is 1.15 (heavy black line), 1.00 (thin black line) or 0.95 (gray line). On the logarithmic vertical axis each geometric schedule is a straight line, so the slope reads directly as the compounding rate. Drawn with PGFPLOTS [32]; values computed with NumPy [33].

7.4 Horizontal depth and state diversity

Let $\mathcal{S}$ be meaningful encounter states and $\pi(s)$ their experienced distribution. The effective state diversity

$$N_{\text{eff}} = \exp\left[-\sum_{s\in\mathcal{S}} \pi(s)\log\pi(s)\right] \tag{7.10}$$

can be much smaller than the number of theoretically possible combinations. If one tool dominates every situation, adding ten inferior tools barely changes experienced diversity.

A useful update adds viable decisions or interactions. Merely adding larger numbers can increase completion time without increasing depth. Track tool choice, route choice, encounter resolution, and repeat behavior to see which of these actually changed.

Practical implication. Design progression around new choices and a readable cadence. Use cost curves to control time, and encounter variation to control depth; do not assume they are interchangeable.

8 Technical reliability as an endogenous retention constraint

Technical reliability changes the experience being measured. A compelling mechanic behind a failed load is not experienced as compelling. A purchase that prompts successfully but fails to deliver changes both revenue and trust. Reliability therefore belongs inside the behavioral model.

Let q_{load} be successful arrival, q_{action} the conditional probability of a working first action, and q_{persist} reliable persistence of earned progress. An illustrative complete-loop probability is

$$q_{\text{loop}} = q_{\text{load}}\, q_{\text{action}}\, q_{\text{persist}}\, q_{\text{design}}. \tag{8.1}$$

The factors need conditional definitions; unconditional probabilities cannot simply be multiplied. This representation highlights complementarity: improving design has little realized value for users who never pass the loading stage.

8.1 Tail performance, not average performance

Let F be frame time in milliseconds. Average FPS is not the inverse of average frame time in general:

$$\mathbb{E}[1000/F] \neq 1000/\mathbb{E}[F]. \tag{8.2}$$

Intermittent stalls matter even when a mean looks acceptable. Report frame-time quantiles, join-time quantiles, crash counts with exposure, and device-specific distributions. Roblox's optimization guidance emphasizes frame rate, memory, and loading, and advises real-device testing [19, 20].

A low FPS percentile is not an average across all players. Comparing a P10 FPS card with a mean FPS target mixes estimands. Likewise, an error count is not the number of affected players: one broken loop can emit thousands of repeated warnings.

8.2 A queueing model of request pressure

For illustration only, consider an M/M/1 queue with arrival rate λ and service rate $\mu > \lambda$. Its mean time in system is

$$W = \frac{1}{\mu - \lambda}, \qquad \rho = \lambda/\mu. \tag{8.3}$$

This is not a claim that Roblox data stores implement M/M/1. It is a useful stress model showing nonlinear sensitivity near saturation:

$$\frac{\partial W}{\partial \lambda} = \frac{1}{(\mu - \lambda)^2}. \tag{8.4}$$

At high utilization, a small request increase can produce a large waiting increase. Retry loops can worsen the effective arrival rate. If each failed attempt triggers an independent retry with probability r, a crude branching approximation gives

$$\lambda_{\text{effective}} = \frac{\lambda_{\text{original}}}{1 - r}, \qquad r < 1. \tag{8.5}$$

A robust system must bound retries and distinguish transient errors from permanent failures. Actual Roblox budgets and limits must come from the current platform documentation, not this illustrative queue [21].

8.3 Reliability-adjusted economics

Suppose the expected retained value of an acquired user conditional on a reliable first loop is L, while unreliable experience yields L_f. With success probability q,

$$\bar{L} = qL + (1 - q)L_f. \tag{8.6}$$

An engineering change increasing q by Δq changes expected value per acquisition by $(L - L_f)\Delta q$. That can justify a reliability fix even when the fix adds no visible feature. The quantities still need estimation; the formula alone does not establish a monetary return.

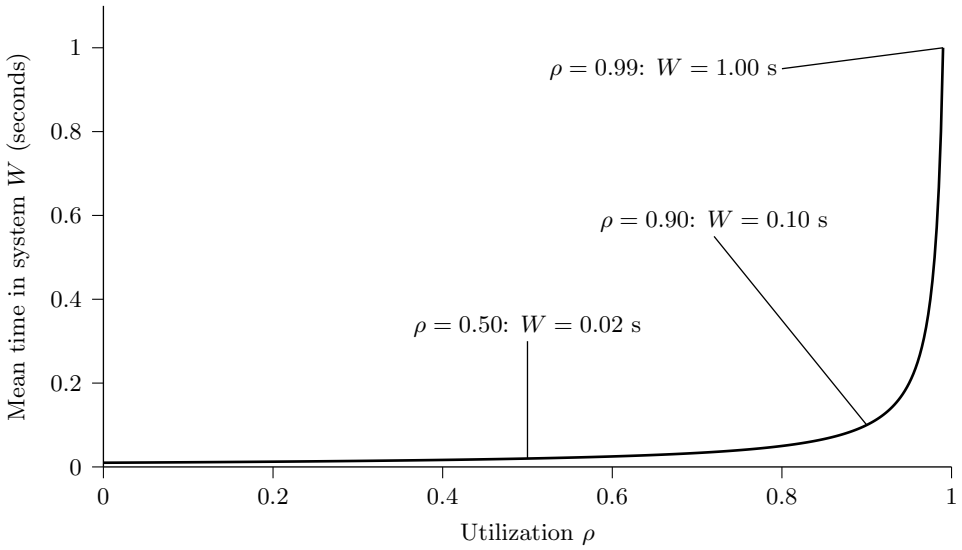


Figure 3. Illustrative M/M/1 sensitivity with service rate 100 requests per second. The curve is $W = 1/[\mu(1-\rho)]$; leader lines mark the three utilizations quoted in the text. This is a queueing stress example, not a measured Roblox service limit. Drawn with PGFPLOTS [32]; values computed with NumPy [33].

8.4 Error prioritization

A useful priority score considers frequency, exposed users, severity, and reproducibility:

$$\begin{aligned} P_j = {} & \Pr(\text{user exposed to error } j)\, \mathbb{E}[\text{loss} \mid \text{exposure } j] \\ & \times \Pr(\text{fix addresses cause} \mid \text{evidence}). \end{aligned} \tag{8.7}$$

Raw log count is an imperfect proxy for the first term and says little about the others. A once-per-purchase receipt failure can matter more than a repeated harmless visual warning.

8.5 A reliability experiment

Deploy a narrowly scoped fix with version tagging. Compare crash-free sessions, successful first loops, persistence verification, and retention across eligible randomized servers where feasible. Confirm that error reductions are not caused by logging removal. For a save change, test earn, disconnect, rejoin, and entitlement reconciliation under load. For a combat change, reproduce object deletion and late-arriving effects.

Practical implication. Treat performance and correctness as measurable product features. Investigate high-impact errors before buying more exposure, and never interpret a falling log count as proof of improved player outcomes without checking the underlying operation.

9 Social play, network effects, and interference

A social game is not a set of independent single-player sessions. Friends affect who joins, whether a server feels alive, what counts as a memorable event, and whether players return together. These effects make community value nonlinear and complicate causal experiments.

Represent potential social connections by adjacency matrix G. For player i, a hypothetical return model is

$$\Pr(Y_{i,t+1} = 1 \mid \mathcal{F}_t) = \text{sigmoid}\left(a_i + b^\top x_{it} + \eta \sum_j G_{ij} Y_{jt} - \zeta\, w_{it} \right), \tag{9.1}$$

where w_{it} is waiting or coordination cost. The coefficient η is not identified by correlation between friends' play: shared preferences, common schedules, and external communication are confounders.

9.1 A branching approximation

Suppose an active player causes an expected k invitations, invitation acceptance is p, and fraction q of accepted users have a meaningful first session. The initial secondary-acquisition factor is

$$R_{\text{invite}} = kpq. \tag{9.2}$$

If descendants behave similarly and there is no overlap or saturation, an initial seed produces expected total reach $1/(1 - R_{\text{invite}})$ when $R_{\text{invite}} < 1$. This is a branching approximation. In real networks, invitations overlap, friends are finite, and acceptance varies. An estimate above one does not prove unlimited growth.

The correct numerator is incremental accepted qualified users, not invitation prompts displayed. Repeatedly prompting the same person can increase event counts while worsening player experience. Logging distinct invite recipients and acceptance paths prevents this accounting error.

9.2 Co-play can improve value without adding users

Even when an invitation produces no new account for the game, coordinating existing players can raise return probability or reduce empty-server waiting. Let $U(n)$ be encounter value at population n. A simple local model is

$$U(n) = a\big(1 - e^{-n/k}\big) - bn^2, \tag{9.3}$$

with positive a, k, b. The first term models the benefit of enough participants; the second models congestion, confusion, or excessive competition. Its derivative

$$U'(n) = \frac{a}{k} e^{-n/k} - 2bn \tag{9.4}$$

can change sign. More players per server is not always better. Matchmaking and server size should support the intended interaction, rather than maximize a population counter.

Roblox's matchmaking documentation describes server selection using player and server characteristics [22]. This manuscript does not assume access to or control over the platform's entire matching algorithm. It treats game-level party handling, mode choice, and encounter capacity as design levers.

9.3 Self-excitation as an update hypothesis

A Hawkes-style intensity model [5] can describe event clustering:

$$\lambda(t) = \mu(t) + \sum_{t_j<t} \kappa e^{-\omega(t-t_j)}. \tag{9.5}$$

Here an observed return or social event increases short-run event intensity. The branching ratio is κ/ω in the scalar model; stationarity requires it below one under standard assumptions. A launch spike driven by a shared external cause can resemble self-excitation. Fitting this equation to aggregated joins does not establish that players caused one another's returns.

A more interpretable experiment randomizes a coordination feature at server or party level and measures group return. The measured lift then concerns the feature under that assignment scheme, not a universal network coefficient.

9.4 Interference changes the estimand

With social spillovers, outcomes depend on the assignment vector:

$$Y_i = Y_i(Z_i, Z_{-i}). \tag{9.6}$$

A player-level A/B test can contaminate itself when friends experience different rules. Server randomization reduces some contamination but may introduce different server populations. Cluster randomization at party or server level should be selected around the actual interaction boundary.

Report direct effects and, where identifiable, spillover effects separately. A feature may benefit treated players partly by shifting costs onto untreated players. Server-wide balance changes especially require attention to total encounter quality.

Practical implication. Build reasons to coordinate, not just reasons to invite. Measure whether social features create better sessions and repeat co-play, and randomize at a unit that respects how players interact.

10 Economy conservation, inflation, and progression integrity

An in-game economy is a flow system. Currency is created, held, spent, transferred, or removed. If measurements do not reconcile these flows, balancing changes can be based on fictional scarcity or fictional abundance.

For player i, let wallet W_i evolve according to earned sources E_i, purchased grants P_i, sinks S_i, and net transfers T_i:

$$W_{i,t+1} = W_{it} + E_{it} + P_{it} - S_{it} + T_{it}. \tag{10.1}$$

For closed player-to-player transfers, $\sum_i T_{it} = 0$. Transfers redistribute currency; they do not create it. Fees or burns are separate sinks. An item trade can transfer value without transferring the game's currency, so item stocks need their own ledger.

10.1 A stock-flow audit

Summing the wallet equation gives

$$\Delta W_{\text{total}} = E_{\text{total}} + P_{\text{total}} - S_{\text{total}} + T_{\text{external}}. \tag{10.2}$$

An unexplained residual

$$\epsilon_{\text{ledger}} = \Delta W_{\text{observed}} - \big(E_{\text{logged}} + P_{\text{logged}} - S_{\text{logged}} + T_{\text{external}}\big) \tag{10.3}$$

may indicate missing logs, duplicates, resets, or a bug. It is not automatically an exploit. The audit should reconcile event ordering, failed transactions, and delayed persistence before assigning a cause.

Roblox economy events support analysis of sources, sinks, and balances; events should reflect completed operations [17, 18]. The analytical ledger remains conceptually separate from the authoritative transaction system. A telemetry event must not be used as the only proof that an entitlement was granted.

10.2 Inflation needs a defined price index

A rising total wallet does not alone establish inflation. More players can hold more currency at unchanged purchasing power. Define a basket of relevant items with prices p_{jt} and fixed weights w_j:

$$I_t = \frac{\sum_j w_j p_{jt}}{\sum_j w_j p_{j0}}. \tag{10.4}$$

If shop prices are fixed but earning rates rise, the more useful quantity may be time-price:

$$\tau_{jt} = \frac{p_{jt}}{y_t}, \tag{10.5}$$

where y_t is an appropriate segment's earning rate. An unchanged 1,000-coin upgrade can become trivial when rewards increase tenfold. Different player segments require different y_t.

10.3 A stable currency sink

A deterministic aggregate approximation is

$$\dot{W} = s - kW, \qquad s, k > 0. \tag{10.6}$$

Its solution is

$$W(t) = W^* + \big(W(0) - W^*\big)e^{-kt}, \qquad W^* = s/k. \tag{10.7}$$

The equilibrium is stable because a larger stock creates larger outflow. This does not imply that proportional currency removal is desirable. The model explains a stability mechanism; actual sinks should create understandable value, such as customization or optional horizontal goals.

When both sources and sinks depend on the stock,

$$\dot{W} = f(W) - g(W), \tag{10.8}$$

a local equilibrium W^* is stable if $f'(W^*) - g'(W^*) < 0$. Multiplicative reward boosts can violate this condition if earning growth outpaces meaningful spending opportunities.

10.4 Item rarity and collection

If an independently earned item has probability p, expected attempts to first acquisition are $1/p$, while

$$\Pr(T > n) = (1-p)^n. \tag{10.9}$$

An expectation conceals the tail. With $p = 0.01$, some players exceed several hundred attempts. A fair design should consider high-percentile effort, duplicate handling, and deterministic alternatives rather than balancing only the mean.

For m equally likely distinct items, the classic coupon-collection expectation is

$$\mathbb{E}[T_{\text{all}}] = m \sum_{k=1}^{m} \frac{1}{k}. \tag{10.10}$$

Unequal rarity makes completion slower and more uneven. Adding many collectible variants can increase apparent content quantity while producing frustratingly low completion probability for casual users.

10.5 Progress preservation

An update changing item strength, sources, or costs also changes the meaning of prior effort. Let V_i^{before} and V_i^{after} be a declared measure of earned progression value. A preservation constraint might require

$$V_i^{\text{after}} \geq \eta \, V_i^{\text{before}} \tag{10.11}$$

for a studio-chosen η. This is a design commitment, not a financial valuation. It encourages migration plans and communication before rebalancing invalidates months of player effort.

Practical implication. Balance sources, sinks, time-prices, and collection tails together. A stable economy should sustain meaningful choices without relying on hidden devaluation or endlessly lengthened waits.

11 Monetization as retained value under uncertainty

Monetization begins with a reliable and valued experience, but engagement does not guarantee purchases. A model must distinguish the probability of paying, the amount paid conditional on paying, refunds or adjustments, delivery correctness, and non-purchase reward streams.

For a cohort and declared horizon, define payer indicator Z_i and positive spend V_i conditional on payment. Then

$$R_i = Z_i V_i, \qquad \mathbb{E}[R_i] = p_{\text{pay}} \, \mathbb{E}[V_i \mid Z_i = 1]. \tag{11.1}$$

The right side is payer conversion times average revenue per payer under matching units and windows. Mixing a daily payer-conversion card with lifetime ARPPU produces an invalid product.

11.1 Heavy tails and small payer counts

Revenue is often concentrated. In a zero-inflated model,

$$\operatorname{Var}(R) = p \operatorname{Var}(V \mid Z = 1) + p(1-p)\{\mathbb{E}[V \mid Z = 1]\}^2. \tag{11.2}$$

A small number of unusually large purchases can dominate uncertainty. A large sample of non-payers does not make conditional spend well estimated when only a handful paid.

Report payer counts, spend distribution, median among payers, upper quantiles where supported, and cohort revenue per acquired user. A bootstrap should resample the assignment unit, preserving a player's purchases together. Resampling individual transactions treats repeated purchases by the same player as independent and can understate uncertainty.

11.2 Discounted cohort value

Let $r(a)$ be probability of activity on day a, $m(a)$ expected net receipts conditional on activity, and $o(a)$ expected operating cost conditional on activity. In a common accounting unit,

$$L = \sum_{a \geq 0} \delta^a r(a)\{m(a) - o(a)\}. \tag{11.3}$$

When monetization depends on purchase history, this factorization must use the appropriate joint expectation rather than multiply unrelated averages. A more general expression is

$$L = \sum_{a \geq 0} \delta^a \mathbb{E}[R_{ia} - O_{ia}]. \tag{11.4}$$

Creator Rewards [28], purchases, and other receipts belong in separately labelled components with their own timing and eligibility. Robux receipts should remain in Robux unless a declared, relevant conversion is applied.

11.3 Pricing and retention externalities

Suppose item demand probability is $q(p)$ at price p, and net unit receipts are $n(p)$. Immediate expected receipts are $n(p)q(p)$. If $n(p) = ap$, a differentiable interior optimum satisfies

$$q(p) + pq'(p) = 0, \tag{11.5}$$

equivalently unit price elasticity. This textbook condition is incomplete when pricing changes future activity:

$$\frac{\mathrm{d}L}{\mathrm{d}p} = \sum_{a} \delta^a \left[\frac{\partial r(a;p)}{\partial p}\, m(a;p) + r(a;p)\, \frac{\partial m(a;p)}{\partial p} \right]. \tag{11.6}$$

An expensive competitive advantage can raise short-run receipts while reducing fair encounters and future player-days. The optimal-price calculation must include that externality or impose a fairness constraint.

11.4 A product is a promise of delivered value

A purchase funnel should distinguish shop exposure, product selection, prompt, completed payment, receipt handling, entitlement delivery, and entitlement persistence. Developer-product receipt handling is an explicit implementation responsibility [23]. A purchase button working visually does not establish the final stages.

The product's description should match delivered behavior. Optional cosmetics, expression, customization, and understandable convenience can create value without making non-paying

players' core experience intentionally defective. This manuscript does not optimize for compulsive purchasing or concealment.

Paid randomized outcomes require current eligibility handling and numerical outcome disclosures under Roblox's documented rules [24]. Those restrictions are product constraints. A formula for expected rarity is not a justification for offering an ineligible transaction.

11.5 Break-even is a distribution, not one number

If acquisition cost per user is c, the relevant decision is not merely whether $\widehat{L} > c$. Under uncertainty, one can require

$$\Pr(L > c \mid \mathcal{D}) \geq 1 - \alpha \tag{11.7}$$

or compare a lower confidence bound with marginal acquisition cost. This is a studio risk rule, not investment advice or a universal threshold. At an early launch, mature L may be mostly unidentified; a sensitivity range is more honest than a precise lifetime estimate.

Practical implication. Resolve purchase reliability, then measure conversion and delivered value by mature cohort. Avoid comparing all historical revenue with only recent advertising spend.

12 Advertising, creative selection, and marginal acquisition

A campaign moves players through several stages: impression, click, arrival, meaningful action, repeat play, and possible purchase. A creative can win one stage while losing the next. The acquisition objective must reflect the desired downstream outcome.

With impressions I, clicks C, qualified new arrivals A, and retained users R under compatible definitions,

$$c_{\text{click}} = C/I, \qquad q_{\text{arrival}} = A/C, \qquad q_{\text{retain}} = R/A, \tag{12.1}$$

$$\frac{R}{I} = c_{\text{click}}\, q_{\text{arrival}}\, q_{\text{retain}}. \tag{12.2}$$

This factorization is an identity only for consistently nested counts. Ads Manager plays, unique acquired users, and dashboard retention cohorts may not be nested in this way. Do not force them into one funnel when attribution windows differ.

12.1 A retained-player objective

For creative j with spend b_j and expected qualified new users $A_j(b_j)$, define

$$\text{CPR}_{j,H} = \frac{b_j}{A_j(b_j)\, r_j(H)}. \tag{12.3}$$

This cost per exact-day retained user differs from cost per player-day,

$$\text{CPPD}_{j,H} = \frac{b_j}{A_j(b_j) \sum_{a=0}^{H} r_j(a)}. \tag{12.4}$$

A lower CTR creative can win either metric if its audience is better matched. Neither metric should be reported as mature before the relevant cohorts have enough follow-up.

12.2 Marginal allocation

Suppose each qualified acquisition from creative j has expected net value L_j in USD. Choose

$$\max_{b_j \geq 0} \sum_j \{L_j A_j(b_j) - b_j\}, \qquad \sum_j b_j \leq B. \tag{12.5}$$

Under concave differentiable A_j, the KKT condition for an active allocation is

$$L_j A_j'(b_j) - 1 = \lambda_B. \tag{12.6}$$

Budget belongs where marginal value is highest, not necessarily where historical average CPP is lowest. Auction saturation and audience fatigue can make average and marginal costs diverge.

For a synthetic response $A_j(b) = a_j(1 - e^{-b/s_j})$, an interior allocation satisfies

$$b_j = s_j \log\left(\frac{L_j a_j}{s_j(1 + \lambda_B)}\right), \tag{12.7}$$

with negative solutions replaced by zero. These parameters require estimation. The formula does not imply that Roblox auctions follow this curve.

12.3 Testing genuinely distinct creatives

Two uploads of the same image are not two independent creative hypotheses. Track visual identity, promise, focal action, composition, and audience. A useful creative matrix varies a small number of clear hypotheses: social tension, immediate action, progression transformation, or a readable risk–reward moment.

Do not change title, onboarding, price, audience, and creative simultaneously and then assign all improvement to the thumbnail. A factorial design can estimate interactions, but requires enough sample in each cell and consistent downstream tracking.

The selection process itself creates winner's curse. If many noisy creatives are tested, the best observed CTR tends to overstate its true rate. Reserve a fresh confirmation sample or use hierarchical shrinkage before assigning the entire budget.

12.4 Stopping and delayed reporting

Roblox documents delayed play and earnings reporting, with data potentially taking 48 hours to appear and attributed activity extending up to 30 days [15]. Those are reporting properties, not proof that continuing ad delivery is required.

If a monitoring process checks spending every Δ hours and delivery rate is bounded by $\bar{b}$ dollars per hour, exposure to unobserved additional spend can be as large as

$$\bar{b}\Delta + \epsilon_{\text{reporting}}. \tag{12.8}$$

Without a reporting-error bound, polling is not a hard spending cap. A campaign stop threshold should reflect the actual control mechanism rather than an assumed billing increment.

Daily and lifetime budgets solve different scheduling problems. Neither is universally more efficient. Compare delivered cohorts and outcomes under matched conditions; do not infer superiority from the budget label alone.

Practical implication. Pay for a declared learning objective or a defensible marginal return. Once the useful contrast is measured, stop new exposure and let existing cohorts mature.

13 Discovery as an uncertain distribution response

Organic distribution is not a deterministic prize for crossing a universal KPI threshold. It is an allocation process matching games to users while other games, audiences, and the platform itself change. A studio observes impressions and resulting behavior, not the complete ranking function.

The current Roblox discovery documentation separates retrieval from ranking and distinguishes users first acquired through Home recommendations from users first acquired elsewhere [13, 14]. Paid traffic may help consideration for retrieval, but its engagement must not be substituted for the Home-acquired cohort's ranking signals. The documentation also presents per-user engagement and return measures; it does not publish a complete coefficient vector allowing creators to calculate rank.

13.1 A reduced-form response model

Let q_t be a vector of measured Home-cohort outcomes, m_t market context, and I_t^{org} organic impressions. A reduced-form planning model is

$$I_{t+1}^{\mathrm{org}} = \Psi_\theta\big(q_{t-\ell:t}, m_t, I_t^{\mathrm{org}}\big) + \epsilon_{t+1}. \tag{13.1}$$

The lag window permits delayed evaluation. Ψ_θ is unknown. A linear regression may describe a local association without identifying the platform's causal decision rule.

If a design intervention u changes both organic response and player behavior, expected future value contains

$$\frac{\mathrm{d}J}{\mathrm{d}u} = \frac{\partial J}{\partial u} + \frac{\partial J}{\partial I^{\mathrm{org}}}\,\frac{\partial \Psi}{\partial q}\,\frac{\partial q}{\partial u}. \tag{13.2}$$

The second term is a possible distribution feedback. Its sign and magnitude cannot be assumed. A robust budget should remain viable under a conservative organic-growth scenario.

13.2 Explore, expand, and audience composition

An initial audience may be unusually well matched. As impressions expand, average conversion can decline because the distribution of users changes. Let z index audience segments:

$$\bar{q}_t = \sum_z w_{zt}\, q_{zt}. \tag{13.3}$$

A falling aggregate can result from changing weights w_{zt} even if every within-segment outcome is stable. Conversely, narrowing delivery can improve aggregate rates without improving the product.

A useful decomposition between two times is

$$\bar{q}_1 - \bar{q}_0 = \sum_z (w_{z1} - w_{z0})\, q_{z0} + \sum_z w_{z1}(q_{z1} - q_{z0}). \tag{13.4}$$

The first term is composition change; the second is within-segment change. Other valid decompositions allocate the interaction differently, so the convention should be reported.

13.3 No universal threshold theorem

A rule such as reaching a particular D1 percentage to guarantee discovery requires evidence that the platform uses that cutoff consistently across games and time. This manuscript has no such evidence. Genre, audience, available alternatives, device experience, and cohort source matter.

Benchmark scorecards can identify areas for investigation, but they are not published causal ranking formulas [10]. A studio should compare relevant cohorts and historical changes rather than optimize a fictional hidden score.

The useful target is player response under the actual exposure distribution. If a feature improves measured first-session experience and repeat play, that is valuable even when organic reach remains unchanged. If reach grows without retained value, the studio should not mistake distribution for a solved product.

13.4 Testing distribution hypotheses

A single before-and-after release is confounded by seasonality and competing events. Where possible, separate a gameplay experiment from a later packaging change. Track organic new cohorts, paid new cohorts, and returning users independently. Preserve the exact date of exposure and the version players actually experienced.

A synthetic control or difference-in-differences comparison can help when no randomization is possible, but only with credible comparison trajectories. An arbitrary popular game is not a valid control merely because it shares a genre. Parallel pretrends, common external conditions, and absence of spillovers require scrutiny.

Practical implication. Improve the game for observed users, then measure whether distribution responds. Do not budget as if an unverified algorithmic multiplier were guaranteed.

14 Causal experiments for games with changing audiences

A feature is useful if it improves relevant outcomes compared with what would otherwise have happened. The counterfactual is difficult because the same player cannot simultaneously experience two openings, two economies, or two match rules.

Let $Z_i \in \{0, 1\}$ be randomized assignment and $Y_i(z)$ the potential outcome. The average treatment effect for the eligible population is

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)]. \tag{14.1}$$

Standard potential-outcome reasoning [1] supports identification when assignment is implemented as intended, outcomes are measured comparably, and the chosen unit adequately handles interference.

14.1 Assignment, exposure, and compliance

An intention-to-treat estimate compares assigned groups:

$$\widehat{\tau}_{\mathrm{ITT}} = \frac{1}{n_1} \sum_{Z_i=1} Y_i - \frac{1}{n_0} \sum_{Z_i=0} Y_i. \tag{14.2}$$

It remains interpretable when some assigned players do not reach the feature. Restricting analysis to those who used it conditions on a post-treatment event and can introduce selection bias.

A trigger-based experiment may enroll a player at the moment a predeclared condition occurs. The trigger must occur before assignment or be unaffected by it. A shop test triggered by shop opening is legitimate if assignment has not already changed whether the player opens the shop.

Record assignment persistently. Re-randomizing each join can expose the same player to conflicting progression rules, contaminate outcomes, and change the estimand from a stable product variant to a sequence of variants.

14.2 Stratification

If device and source strongly predict outcomes, randomize within strata z and estimate

$$\widehat{\tau} = \sum_{z} w_z \, \widehat{\tau}_z. \tag{14.3}$$

The weights w_z should represent a declared target population. Reporting only favorable strata after looking at results creates a multiple-testing problem. Prespecify important heterogeneity contrasts and retain the aggregate effect.

A mobile improvement may be masked by desktop-heavy traffic. Conversely, a desktop gain may not justify a rollout to a mobile-dominated launch. Stratification improves both interpretability and, often, precision.

14.3 Sample size and minimum useful effect

For a rough two-arm binary-outcome calculation with equal group sizes,

$$n_{\text{arm}} \approx \frac{2\bar{p}(1-\bar{p})\,(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}. \tag{14.4}$$

This normal approximation assumes independent units and a fixed planned analysis. It is a planning aid, not an exact test. Detecting a change from 5% to 6% with conventional two-sided 5% significance and 80% power requires approximately 8,160 users per arm under this approximation.

The minimum detectable effect should be economically or experientially meaningful. If only a tiny effect justifies the development cost, a small launch test may not resolve the decision. Increasing the number of variants then makes the problem worse by splitting the sample.

For equal-size clusters of mean size m with intracluster correlation ρ, a rough design effect is

$$\text{DE} = 1 + (m-1)\rho. \tag{14.5}$$

Multiply the independent-unit sample estimate by this factor as a first approximation. Actual uneven cluster sizes and time dependence require better analysis.

14.4 Guardrails and stopping rules

Choose one primary estimand, a small set of guardrails, and a fixed or valid sequential decision rule. Relevant guardrails include critical failures, purchase delivery, fairness, and device-specific degradation. A session-time increase caused by a longer unskippable wait should not be accepted as improved engagement.

Prespecify what ends the test: a sample size, follow-up horizon, maximum duration, safety failure, or statistically valid boundary. Repeatedly checking ordinary confidence intervals and stopping at the first favorable result inflates false discoveries.

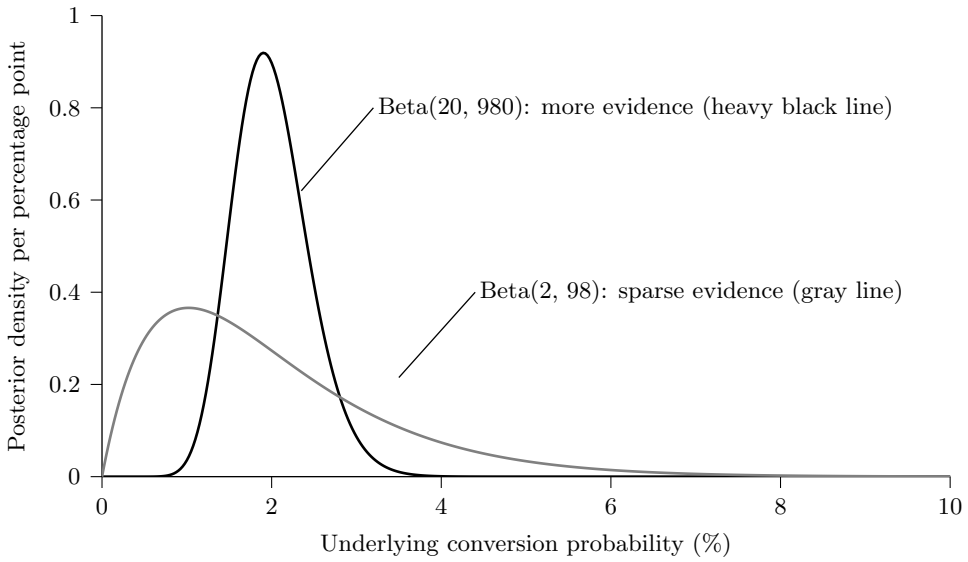


Figure 4. Two synthetic Beta posteriors with the same mean, 2 percent, and different concentration: Beta(20, 980) (heavy black line) and Beta(2, 98) (gray line). The curves illustrate uncertainty, not observed conversion rates for a game. Drawn with PGFPLOTS [32]; densities computed with SciPy [34].

14.5 Interference-aware release units

Balance changes within shared rounds often require server-level assignment. Social coordination features may require party-level assignment. Switchback designs alternate treatment by time blocks but must account for carryover, weekday effects, and shared persistent economies.

Practical implication. A good experiment makes one consequential uncertainty smaller. It need not test every feature simultaneously, and it should not be judged by whether it produces a favorable headline.

15 Statistical inference, shrinkage, and sequential decisions

A point estimate is a compressed summary of uncertainty. At small counts it can be misleading: zero observed payers is not proof of zero payer probability, and a high D1 estimate from a tiny cohort can be mostly noise.

For k successes in n independent Bernoulli trials, likelihood is

$$L(p) \propto p^k (1-p)^{n-k}. \tag{15.1}$$

With $\mathrm{Beta}(a, b)$ prior, the posterior is

$$p \mid \mathcal{D} \sim \mathrm{Beta}(a+k,\, b+n-k), \qquad \mathbb{E}[p \mid \mathcal{D}] = \frac{a+k}{a+b+n}. \tag{15.2}$$

The prior should be declared and sensitivity checked. A credible interval is conditional on the prior and model; it is not automatically a frequentist confidence interval.

15.1 A score interval for a binary rate

The Wilson interval uses

$$\frac{\widehat{p} + z^2/(2n) \pm z\sqrt{\widehat{p}(1-\widehat{p})/n + z^2/(4n^2)}}{1 + z^2/n}. \tag{15.3}$$

It behaves better near boundaries than the naive Wald interval. Independence, a well-defined denominator, and adequate follow-up remain necessary. Reporting an interval for an unmatured cohort does not repair its missing future outcomes.

If $k = 0$, the approximate one-sided 95% upper bound is about $3/n$ for moderate n, from $(1-p)^n \approx e^{-np}$. A zero count in 300 users is consistent with nontrivial rare conversion. The approximation should not replace an exact calculation when precision matters.

15.2 Hierarchical partial pooling

For creatives j, a logistic random-effects model is

$$k_j \sim \text{Binomial}(n_j, p_j), \qquad \text{logit}(p_j) = \mu + x_j^\top \beta + u_j, \qquad u_j \sim N(0, \sigma_u^2). \tag{15.4}$$

Sparse creatives shrink toward a shared distribution while well-measured creatives retain more individual information. The model can reduce winner's curse, but it can also overshrink genuinely different populations if the hierarchy is inappropriate.

The design matrix should represent meaningful creative distinctions, not arbitrary filenames. Source, date, placement, and audience covariates help describe comparability. They do not fully eliminate nonrandom delivery bias.

15.3 An elementary time-uniform bound

For independent Bernoulli observations with common mean p, Hoeffding's inequality gives

$$\Pr(|\widehat{p}_n - p| > \epsilon) \le 2e^{-2n\epsilon^2}. \tag{15.5}$$

Allocate error $\alpha_n = \alpha/[n(n+1)]$, whose sum is α. Setting

$$\epsilon_n = \sqrt{\frac{\log\left(2n(n+1)/\alpha\right)}{2n}} \tag{15.6}$$

and taking a union bound yields

$$\Pr\{\forall n \ge 1 : |\widehat{p}_n - p| \le \epsilon_n\} \ge 1 - \alpha. \tag{15.7}$$

Clip the interval to $[0, 1]$. This conservative construction is an explicit, valid alternative to repeated fixed-time intervals under its stated assumptions. More efficient confidence-sequence methods are developed by Howard and colleagues [6].

Adaptive assignment, clustered outcomes, and delayed rewards require corresponding martingale or design-aware methods; the simple IID bound does not automatically cover them. For two independent arms, allocate error to both sequences and subtract their interval endpoints to bound the difference.

15.4 Off-policy evaluation

When a policy chooses action A_i with known propensity $e(A_i \mid X_i)$, inverse-propensity weighting can estimate another policy's value:

$$\widehat{V}_{\mathrm{IPS}}(\pi) = \frac{1}{n} \sum_i \frac{\pi(A_i \mid X_i)}{e(A_i \mid X_i)} Y_i. \tag{15.8}$$

Overlap is essential. Tiny propensities create high variance; zero propensity makes the corresponding action unidentifiable. Ads Manager typically does not expose every user-level delivery propensity needed for this estimator, so it should not be applied to aggregate campaign rows as if those inputs existed.

With a reward model $\widehat{m}$, the doubly robust form is

$$\begin{aligned} \widehat{V}_{\mathrm{DR}}(\pi) = \frac{1}{n} \sum_i \Bigg[& \sum_a \pi(a \mid X_i)\, \widehat{m}(X_i, a) \\ & + \frac{\pi(A_i \mid X_i)}{e(A_i \mid X_i)} \{Y_i - \widehat{m}(X_i, A_i)\} \Bigg]. \end{aligned} \tag{15.9}$$

Dudík and colleagues develop this approach [7]. Cross-fitting helps avoid overfitting nuisance estimates. Double robustness does not cure hidden confounding or absent support.

Practical implication. Use uncertainty to decide whether to act, wait, or run a sharper test. Do not use a complex estimator when its required assignment or follow-up data do not exist.

16 Maturity, missingness, and delayed outcome accounting

A launch dataset is triangular. On extraction day T, a cohort arriving at c has follow-up only through age $T - c$. Later outcomes are unavailable, not negative.

Let $Y_{ic}(a)$ be an outcome at cohort age a, and

$$O_{ic}(a; T) = \mathbf{1}\{c + a \le T, \text{ event processed by } T\}. \tag{16.1}$$

A valid exact-day estimate uses cohorts with enough calendar follow-up and adequate processing maturity. Dividing observed day-30 returns by all recent acquisitions implicitly treats future days as failures.

16.1 A reporting-latency model

Let $F_\ell(s)$ be the probability an event is processed within s time units. For events generated at times t_e, expected observed count is

$$\mathbb{E}[N_{\mathrm{observed}}(T)] = \sum_e \Pr(\text{event } e)\, F_\ell(T - t_e). \tag{16.2}$$

Recent events receive smaller observation probability. Nowcasting divides by or models this probability, but it can be unstable when F_ℓ is small or changes over time. A short waiting buffer is often more defensible than a highly parameterized correction.

Roblox's documented reporting delay is an operational reason to preserve successive snapshots [15]. A revised count can reflect verification or attribution processing, not newly purchased exposure. Record both extraction time and the reported window.

16.2 Missingness mechanisms

Missing completely at random, missing at random conditional on observed variables, and missing not at random lead to different estimators. A device crash that prevents its own telemetry from uploading is a plausible nonrandom missingness mechanism. The observed sessions can then look healthier than the experienced sessions.

If observation propensity $\pi_i = \Pr(O_i = 1 \mid X_i)$ is known or consistently estimated under conditional ignorability, a weighted mean is

$$\widehat{\mu} = \frac{\sum_i O_i Y_i / \pi_i}{\sum_i O_i / \pi_i}. \tag{16.3}$$

This does not identify unobserved outcomes when missingness depends on unmeasured severity. Sensitivity bounds should accompany strong assumptions.

For a binary outcome with n_o observed successes k and n_m missing outcomes among $n = n_o + n_m$, assumption-free bounds are

$$\frac{k}{n} \le p \le \frac{k + n_m}{n}. \tag{16.4}$$

Wide bounds accurately express missing information. Replacing missing values with zero produces one extreme, not an unbiased estimate.

16.3 Bounding immature value

Suppose per-user net daily receipts are bounded above by $\bar{v}$ and nonnegative. After observing through age H, the unobserved discounted tail is bounded by

$$0 \le L - L_H \le \frac{\bar{v}\,\delta^{H+1}}{1-\delta}, \qquad \delta < 1. \tag{16.5}$$

This bound may be very loose. If an additional defensible bound $r(a) \le C r^a$ holds with $0 < r < 1$,

$$L - L_H \le \frac{\bar{v} C (\delta r)^{H+1}}{1 - \delta r}. \tag{16.6}$$

The decay assumption must be validated or presented as a scenario. It is not legitimate to fit an early decline and proclaim a fully known lifetime value.

16.4 A practical extraction calendar

Maintain an immediate operational snapshot, a processing-stable short-horizon snapshot, mature D7 cohorts, and a final planned attribution-horizon snapshot. Preserve the same campaign identifiers and cohort definitions throughout. Use custom date ranges covering the complete launch when a rolling window would exclude its beginning.

A 28-day default window cannot fully represent a 30-day post-acquisition tail plus processing time. At final extraction, record both the launch interval and follow-up interval explicitly. This is not permission to select a flattering range; it is a requirement to cover all relevant exposure and outcomes.

Practical implication. Freeze spending independently of the evidence clock. Let the existing players generate return and purchase outcomes, then extract with cohort age, processing maturity, and complete launch coverage visible.

17 Production capacity, technical debt, and the marginal feature

The studio's scarce resource is not only advertising cash. Engineering attention, art capacity, testing time, and the ability to maintain shipped systems constrain what can be sustained. A feature creating future support work consumes capacity beyond its release week.

Let h_j be effort allocated to improvement j and $v_j(h_j)$ its expected contribution to a declared objective. A local resource-allocation problem is

$$\max_{h_j \geq 0} \sum_j v_j(h_j), \qquad \sum_j h_j \leq H. \tag{17.1}$$

If each v_j is concave and differentiable, active projects satisfy

$$v_j'(h_j) = \lambda_H. \tag{17.2}$$

λ_H is the shadow value of an additional development hour under the model. A large visible feature with low marginal value can be dominated by a small save-load fix.

17.1 Complementarity changes prioritization

Independent feature values are often false. If onboarding quality x and reliability y combine multiplicatively,

$$V(x, y) = K(1 - e^{-ax})(1 - e^{-by}), \tag{17.3}$$

then

$$\frac{\partial^2 V}{\partial x \, \partial y} = Kab\, e^{-ax-by} > 0. \tag{17.4}$$

Improving one increases the value of the other. A roadmap ranked by isolated estimates misses this interaction. Conversely, two cosmetic features competing for the same player attention can have negative interaction.

A portfolio experiment should identify prerequisites. A polished upgrade screen cannot demonstrate value when the preceding reward never saves. Shipping both together may be operationally necessary, but the evidence should then concern the bundle rather than an invented separate effect.

17.2 Technical debt as a state variable

Let d_t denote maintenance burden. A simple dynamic is

$$d_{t+1} = (1 - \gamma)d_t + \eta f_t - \kappa m_t, \tag{17.5}$$

where f_t is new feature throughput and m_t maintenance effort. Enforce $d_t \geq 0$ through a nonnegative transition rule. Available future capacity might be $H_t^{\text{effective}} = H_t - \phi(d_t)$ with increasing ϕ.

The equation is a planning model. Its coefficients should be inferred from actual regressions, incident time, and release delays. It explains why a rapid feature week can reduce later throughput even when its immediate player response is favorable.

A feature's net value should include maintenance:

$$V_j^{\text{net}} = V_j^{\text{player}} - C_j^{\text{build}} - \sum_{t > t_j} \delta^{t-t_j} C_{jt}^{\text{maintenance}}. \tag{17.6}$$

Reusable tools may have modest first-release payoff but improve many future releases. Their benefit belongs to the future workflow they enable, not to a vague claim that infrastructure always pays for itself.

17.3 Dependencies and the critical path

If tasks form a directed acyclic graph, task finish times satisfy

$$C_j = d_j + \max_{i \in \mathrm{pred}(j)} C_i. \tag{17.7}$$

The release date is the maximum finish time of required tasks. Adding an artist to a task outside the critical path may not accelerate release. An uncertain integration task on the critical path deserves early validation.

Stochastic durations require a distribution of release dates. Summing only average durations understates risk when a release waits for the maximum of several uncertain tasks. A simple Monte Carlo plan can expose this bottleneck without pretending to know exact future completion times.

17.4 A staged production policy

Build the smallest testable loop, validate correctness, instrument its consequential steps, and test player understanding. Only then expand content breadth. This is not a universal instruction to make every game tiny. It is a way to discover whether the central interaction works before its surrounding systems become expensive to change.

The relevant question for a proposed feature is: which transition, choice, or recurring value does it improve? If the answer is only that comparable games have it, the roadmap has not identified its mechanism.

Practical implication. Allocate hours to the strongest constrained marginal improvement. Include dependencies and maintenance, and distinguish a feature's appeal in a design meeting from its demonstrated effect in play.

18 A unified constrained control formulation

The preceding models can be assembled into one objective without claiming a closed-form recipe for success. The purpose of the larger formula is to expose dependencies, units, and unknowns, not to replace observation.

Let q_t be product quality and content state, d_t maintenance burden, b_{jt} advertising spend by creative, and h_{kt} production effort. Let A_{cz} be new users acquired at date c in segment z, with acquisition response

$$A_{tz} = \mathcal{A}_z(b_t, q_t, \mathrm{market}_t; \theta) + \xi_{tz}. \tag{18.1}$$

For cohort transitions $P_{a,t,z}$, define a daily activity kernel

$$r_{cz}(a) = e_A^\top \left(\prod_{s=0}^{a-1} P_{s,c+s,z}(q_{c+s}, u_{c+s}; \theta) \right) e_A. \tag{18.2}$$

At $a = 0$ the empty product is the identity. This representation uses a daily active-state definition; it does not equate instantaneous occupancy with daily participation.

18.1 The integrated objective

With expected net receipts per active user μ_{czt}, a finite-horizon formulation is

$$\begin{aligned}\max_{\pi} \quad \mathbb{E}_{\pi} \sum_{t=0}^{T} \delta^t \Bigg[& \sum_{z} \sum_{c=0}^{t} A_{cz} r_{cz}(t-c) \mu_{czt} \\ & - \sum_{j} b_{jt} - C_{\mathrm{dev}}(h_t) - C_{\mathrm{ops}}(X_t) + \omega V_{\mathrm{player}}(X_t, u_t) \Bigg] - \lambda \mathcal{R}(\pi).\end{aligned} \tag{18.3}$$

Subject to

$$q_{t+1} = \mathcal{Q}(q_t, h_t, d_t, \xi_t), \tag{18.4}$$

$$d_{t+1} = \mathcal{D}(d_t, h_t), \tag{18.5}$$

$$\sum_{k} h_{kt} \le H_t, \qquad \sum_{t,j} b_{jt} \le B, \tag{18.6}$$

$$\Pr(\text{critical failure} \mid z, t) \le \epsilon_z. \tag{18.7}$$

All monetary terms require one accounting unit. μ can be negative after operating costs, and source-specific policy restrictions remain feasibility constraints.

Equation (18.3) is a synthesis, not an empirically validated universal law. Unknown acquisition functions, player transitions, and market disturbances make a literal global solution unrealistic. Its practical use is to identify which uncertain marginal quantity controls the next decision.

18.2 Bellman structure

Following the dynamic-programming perspective [9], for a fully observed state x with known transition law, the finite-horizon value recursion is

$$V_t(x) = \max_{u \in \mathcal{U}_t(x)} \big\{ r_t(x,u) + \delta\, \mathbb{E}[V_{t+1}(X_{t+1}) \mid x, u] \big\}. \tag{18.8}$$

Under partial observation, the state must include a belief distribution over latent variables and parameters, or a justified approximation to that belief. Substituting an unobserved true state into the recursion is not an implementable policy.

A feature test has dual value: it may improve play now and teach the studio which future action to take. This exploration value should be considered when a small experiment can prevent a large mistaken release. It does not justify testing without limits.

18.3 Adjoint sensitivity for a smooth approximation

For deterministic dynamics $\dot{x} = F(x,u)$ and objective

$$J = \int_0^T e^{-\rho t} \ell(x,u)\, \mathrm{d}t + \Phi(x(T)), \tag{18.9}$$

define the present-value Hamiltonian

$$\mathcal{H} = e^{-\rho t} \ell(x,u) + \lambda^{\top} F(x,u). \tag{18.10}$$

The adjoint satisfies

$$\dot{\lambda} = -e^{-\rho t}\ell_x - F_x^\top \lambda, \qquad \lambda(T) = \Phi_x(x(T)). \tag{18.11}$$

The control gradient is

$$\frac{\delta J}{\delta u(t)} = e^{-\rho t}\ell_u + F_u^\top \lambda. \tag{18.12}$$

This equation propagates downstream value back to an earlier development action. It generalizes the first-session hazard derivative: an early improvement can affect many later states.

Nonconvex objectives can have multiple local optima [8]. A gradient method finding one stationary point is not proof of the best possible game. Discrete content decisions, multiplayer interference, and uncertain parameters further limit literal application.

18.4 Robustness and downside control

For plausible parameter set Θ, robust planning solves

$$\max_{\pi} \min_{\theta \in \Theta} J(\pi; \theta). \tag{18.13}$$

A less pessimistic alternative penalizes downside loss using

$$\mathrm{CVaR}_\alpha(L) = \min_{\eta} \left\{ \eta + \frac{1}{1-\alpha}\mathbb{E}[(L-\eta)_+] \right\}. \tag{18.14}$$

Here α is the tail level and L is loss, not player lifetime. These are risk preferences chosen by the studio. They do not turn uncertain forecasts into guarantees.

Practical implication. Use the unified formulation as a map of dependencies. Estimate local effects, impose hard operational constraints, and update the policy as evidence arrives.

19 Synthetic experiments and falsifiable predictions

The following experiments are synthetic. They do not estimate Roblox's ranking algorithm, fit *Murder Mystery 2* telemetry, or forecast a particular game's revenue. Their purpose is to verify model implications and reveal which assumptions matter. The numerical computations use deterministic equations and explicitly specified parameters.

19.1 Two first-session interventions

Take a baseline hazard of 0.8 per minute during the first minute and 0.32 thereafter. Compare an early intervention reducing the first-minute hazard to 0.5 with a late intervention reducing hazard from 0.32 to 0.20 after minute 3. For each,

$$S(t) = e^{-H(t)}, \qquad H(t) = \int_0^t h(s)\,\mathrm{d}s. \tag{19.1}$$

The baseline has $S(5) = e^{-2.08} \approx 0.1249$. The early change gives $e^{-1.78} \approx 0.1686$; the late change gives $e^{-1.84} \approx 0.1588$. The interventions remove different total hazards, so these numbers are not an equal-cost comparison.

For equal small reductions in cumulative hazard localized at different times, equation (5.8) predicts a larger restricted-mean gain earlier when value weights are nonnegative. The experiment can be falsified in a real game if the intervention changes downstream value or later behavior in ways omitted by the model.

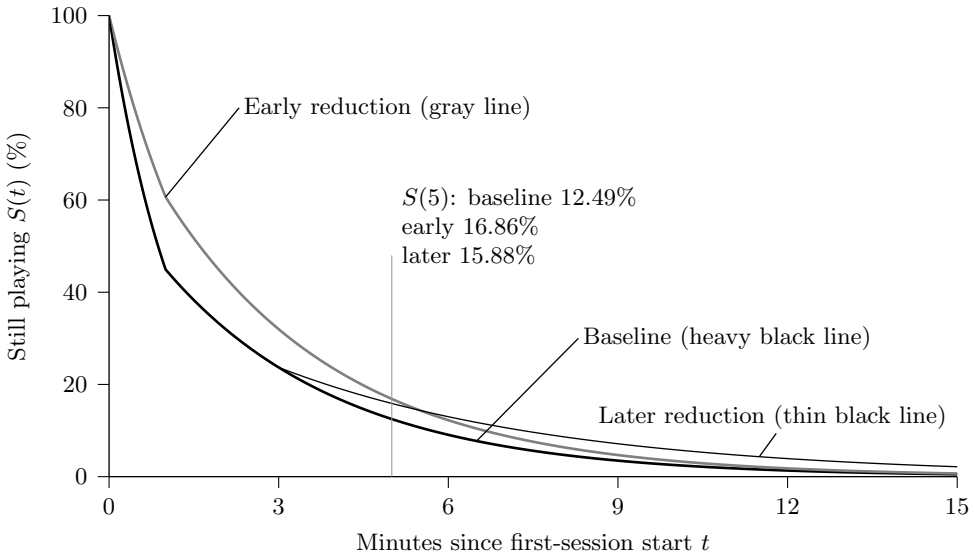


Figure 5. Synthetic first-session survival $S(t) = e^{-H(t)}$ under the piecewise hazards specified in the text: baseline (heavy black line), early reduction (gray line) and later reduction (thin black line). The later-reduction curve coincides with the baseline until minute 3, where its hazard first differs; the thin vertical rule marks $t = 5$, where the three values quoted in the text are read. Curves illustrate model consequences, not observed treatment effects. Drawn with PGFPLOTS [32]; values computed with NumPy [33].

19.2 Acquisition pulse versus return kernel

Simulate 120 days with 1,000 new users daily through day 6 and no new users afterward. Compare

$$r_{\text{weak}}(0) = 1, \qquad r_{\text{weak}}(a) = 0.04\, e^{-0.18(a-1)},\ a \geq 1, \tag{19.2}$$

$$r_{\text{strong}}(0) = 1, \qquad r_{\text{strong}}(a) = 0.20\, e^{-0.06(a-1)},\ a \geq 1. \tag{19.3}$$

Convolution with the same acquisition pulse yields markedly different tails. The model contains no organic feedback. Any real post-stop audience increase would require additional acquisition, reactivation, changing kernels, or measurement revisions.

19.3 A queue sensitivity experiment

Set the illustrative service rate to $\mu = 100$ requests per second and vary utilization ρ. The M/M/1 formula gives $W = 1/[100(1 - \rho)]$ seconds. At 50% utilization $W = 0.02$; at 90%, 0.10; at 99%, 1.00. This demonstrates sensitivity near saturation. It is not a Roblox data-store capacity estimate.

The falsifiable operational prediction is that reducing duplicate or unnecessary requests should reduce a measured bottleneck under a controlled load test, if request pressure is its cause. If latency remains unchanged, investigate another bottleneck rather than force the queue explanation.

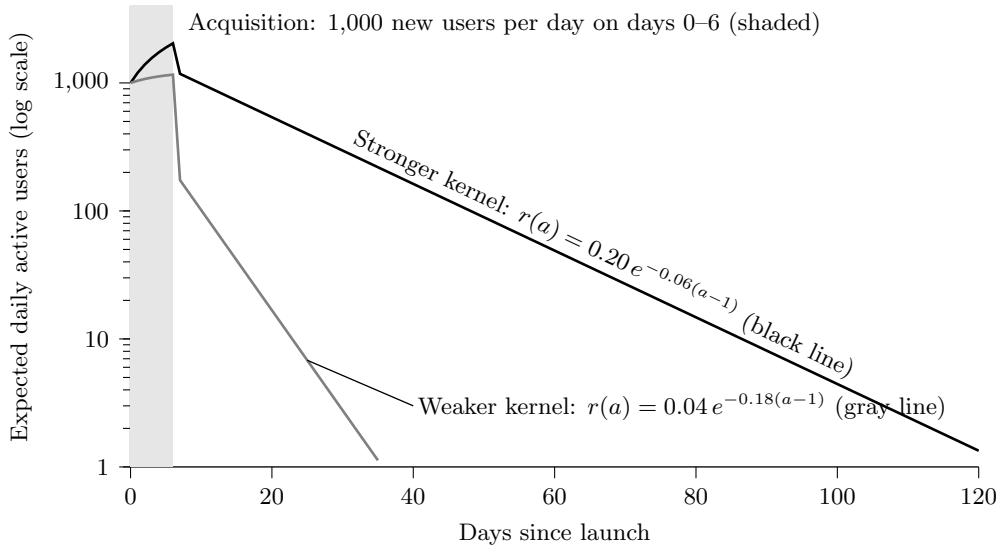


Figure 6. Synthetic cohort accounting under identical acquisition: stronger return kernel (black line) and weaker return kernel (gray line). The shaded period is the seven-day acquisition pulse. The vertical axis is logarithmic, so after the pulse each exponentially decaying tail appears as a straight line whose slope is the kernel's decay rate; the weaker kernel leaves the plotted range when its expected population falls below one user. The longer tail comes from the return kernel, not further spending. Drawn with PGFPLOTS [32]; convolution computed with NumPy [33].

19.4 Creative allocation and uncertainty

Use two synthetic acquisition responses:

$$A_1(b) = 6000\big(1 - e^{-b/100}\big), \qquad A_2(b) = 4000\big(1 - e^{-b/80}\big). \tag{19.4}$$

Assign hypothetical net per-user values $L_1 = 0.02$ and $L_2 = 0.04$ USD. With a \$100 total budget, optimize the concave objective from Section 12. Neither the CTR nor the initial acquisition slope alone identifies the best allocation; retained value and saturation matter jointly.

19.5 Sensitivity before precision

Repeat each planning calculation over ranges of departure rates, net user value, development cost, and update impact. If the recommended action changes across plausible values, report the decision boundary. A narrow point forecast is not more useful than a broad sensitivity analysis when the parameters are poorly identified.

Simulation should also test accounting invariants: population mass, nonnegative wallets, stochastic transition matrices, unique receipts, and consistent cohort denominators. A model that violates these invariants should not be calibrated to player data.

Practical implication. Use simulations to expose assumptions and design a discriminating test. Their purpose is not to decorate a forecast with scientific-looking curves.

20 A preliminary launch audit: Push Party

This case uses an author-authorized Creator Dashboard and Ads Manager extraction from October 26, 2026 [30]. It is a descriptive, early-launch snapshot, not a randomized study. Several

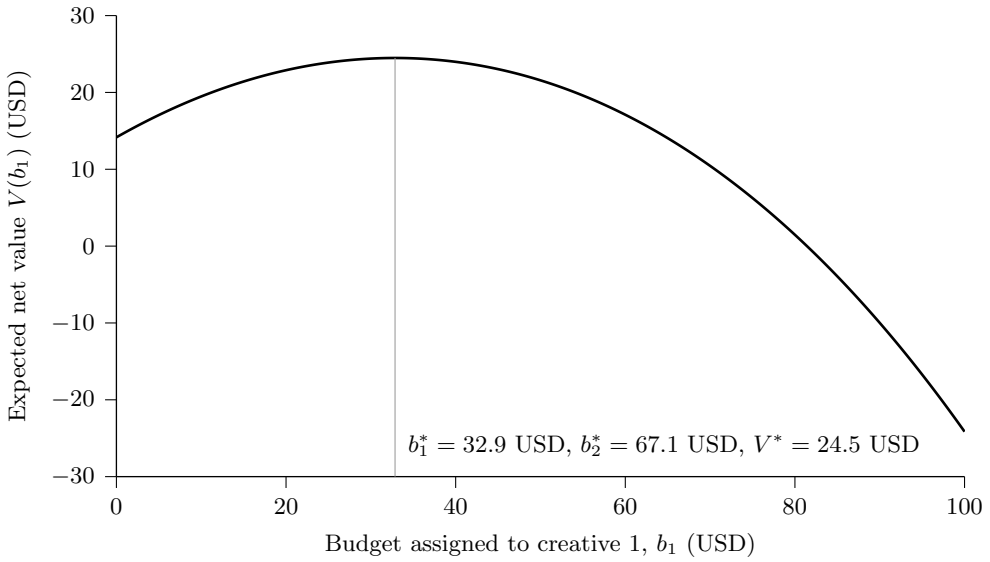


Figure 7. Synthetic net value $V(b_1) = L_1 A_1(b_1) + L_2 A_2(100 - b_1) - 100$ along a fixed \$100 budget split; the thin vertical rule marks the maximizer. The maximizer is conditional on assumed acquisition curves and user values. It is not a recommended real campaign split. Drawn with PGFPLOTS [32]; maximizer found with SciPy [34].

endpoints update asynchronously. The evidence package contains 33 dashboard captures, five campaign rows, 24 creative records, and 500 grouped error-report entries.

The case is valuable because apparently favorable acquisition can coexist with weak return behavior, incomplete attribution, and reliability concerns. It should not be used as a universal benchmark for another genre.

20.1 Observed acquisition and engagement

The 6.7-minute average-playtime card was an average over days including tiny prelaunch samples. Busy raw daily rows showed roughly 2.4–3.4 minutes. This is a concrete example of the average-of-ratios issue, not evidence that either dashboard necessarily malfunctioned.

20.2 Cohort differences

A mandatory tutorial had already been made optional. The smaller newer cohort improved, but the newest large cohort did not yet have mature D1 evidence. Without the exact deployment time, consistent source composition, and mature comparison, attributing the change to the tutorial fix would be premature.

Some Explore daily labels appeared one day earlier than UTC cohort and Overview labels. The extraction preserved labels rather than silently shifting them. An analyst should resolve this alignment before joining the views.

20.3 Campaign comparison

The white-character campaigns had higher observed CTR. They were not randomized against the others under identical conditions. The lowest reported CPP belonged to another campaign.

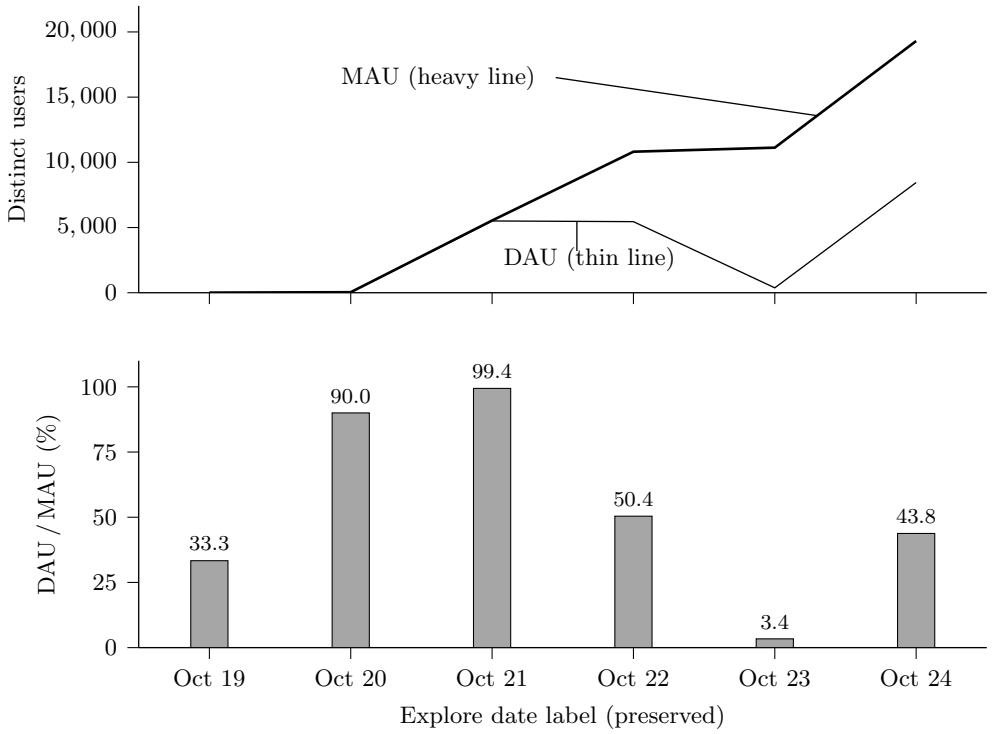


Figure 8. Observed Explore daily counts and stickiness from the provisional audit. Top: distinct monthly (heavy line) and daily (thin line) active users as rendered. Bottom: the ratio DAU/MAU in percent, printed above each bar. Date labels are preserved exactly; they must not be silently aligned with the UTC cohort table. High launch stickiness can coexist with weak repeat play. Drawn with PGFPLOTS [32] from the audit values as rendered.

Thus the evidence supports a creative hypothesis, not a causal ranking by retained value.

20.4 Reliability and observability

The Last 1 day error report showed 114,220 occurrences, including 39,634 queued datastore warnings, 22,133 throttled GetAsync errors, 4,591 combat-effects nil-parent errors, and 719 ProcessReceipt callback-access errors. Five crate configurations each had 588 missing-price/product-ID warnings. These are occurrence counts, not affected-user counts; first-seen dates were historical.

The logs identify investigation priorities. They do not prove that all purchases failed, that data was lost, or how much retention each issue caused. Funnel and economy dashboards had no configured events, limiting the ability to locate progression and purchase abandonment precisely.

20.5 A stopping decision with an open observation window

At a later check, all five campaign switches were off, with two campaigns labelled Rejected and three Paused. Ads Manager said the experience was suspended for ad serving, while the game remained Public and its younger-audience review had passed. The underlying moderation reason was not exposed. This status contradiction is not evidence that every creative independently failed.

The rational decision was to stop new paid exposure, preserve the acquired cohorts, inves-

Table 3. Selected observed launch metrics. Windows and denominators differ; rows must not be multiplied indiscriminately.

Measure	Observed value and qualification
New acquired users	19,234 in acquisition view
Sponsored acquisitions	18,655, approximately 96.99%
Sponsored D1	1.43%, source-cohort measure
New-session survival at 5 minutes	12.43%
Sponsored first-day playtime	2.9 minutes per acquired user, cumulative
DAU	2,832 daily-average card
Stickiness	47.34% daily-average DAU/MAU
Home recommendation plays	0 in the inspected source table
Net purchase receipts	53 Robux, Last 7 days
Creator Rewards	170 Robux in a separately updated Last 7 days widget

Table 4. Retention percentages as rendered. Rounded percentages must not be reverse-engineered into exact success counts.

UTC cohort	New users	D1	D2	D3
Oct 24	306	7.82%	N/A	N/A
Oct 23	5,330	0.90%	1.18%	N/A
Oct 22	5,459	1.53%	0.24%	1.31%
Oct 21	30	6.67%	0.00%	3.33%

tigate reliability, and wait for mature outcomes. This decision did not require declaring final lifetime value. It followed from the available traffic volume, weak observed early engagement, unresolved serving eligibility, and the opportunity cost of further spending.

Practical implication. Transfer the measurement lessons to a new game. Do not transfer a preliminary CTR winner as though it established that game's audience fit or economics.

21 Analytical limits and useful propositions

A sophisticated blueprint should state what its inputs cannot establish. Several simple results prevent surprisingly consequential mistakes. The propositions in this section are elementary consequences of the specified models; they are not claims of new general mathematical theory.

21.1 One retention point does not identify lifetime value

Proposition 21.1 (Nonidentification from D1)**.** *Knowing $r(0) = 1$ and $r(1) = p$ does not identify $L_H = \sum_{a=0}^{H} r(a)$ for $H \geq 2$, even if $r(a)$ is nonincreasing.*

Proof. One compatible kernel has $r(a) = 0$ for $a \geq 2$, giving $L_H = 1 + p$. Another has $r(a) = p$ for $1 \leq a \leq H$, giving $L_H = 1 + Hp$. Both are nonincreasing and share the observed D1. For $p > 0$ and $H > 1$, the values differ. □

Table 5. Sequentially captured Last 28 days campaign rows. Rounded rows total $474.72 versus a contemporaneous large spend metric of $474.70; a later refresh showed $474.80.

Campaign label	Spend USD	CTR	Reported CPP
White Character Relaunch	39.26	4.5%	0.056
Proven White Creative	129.43	4.7%	0.023
Test Creatives	56.42	1.7%	0.032
October 22 launch	91.97	2.0%	0.018
October 21 launch	157.64	2.3%	0.021

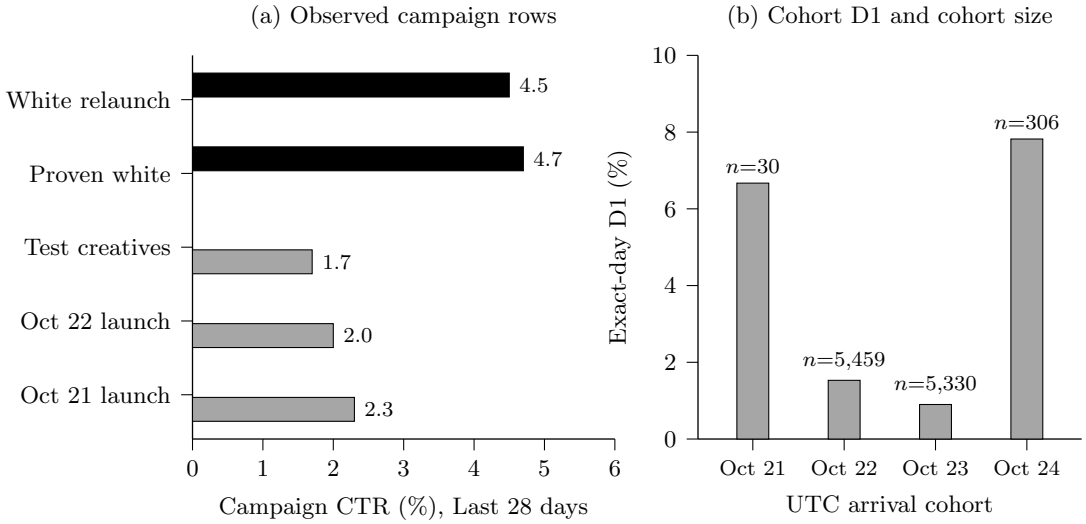


Figure 9. Observed CTR and cohort D1 from different views, as rendered in Tables 4 and 5. (a) Campaign click-through rates; black bars are the two white-character campaigns, gray bars the others. Campaigns were not randomized. (b) Exact-day D1 by UTC arrival cohort with cohort size n above each bar; the two cohorts with higher D1 are far smaller than the two large launch cohorts. Drawn with PGFPLOTS [32] from the values in Tables 4 and 5.

Exact-day retention is not generally nonincreasing, so the actual identification problem is weaker still. A strong D1 result is informative, but cannot supply a mature D30 or lifetime revenue estimate by itself.

21.2 A value-of-information bound

Suppose actions $a \in \mathcal{A}$ have uncertain values $V(a, \theta)$. With current data D, the best expected value is $\max_a \mathbb{E}[V(a, \theta) \mid D]$. Perfect information would give $\mathbb{E}[\max_a V(a, \theta) \mid D]$. Their difference is

$$\mathrm{EVPI} = \mathbb{E}\Big[\max_a V(a, \theta) \;\Big|\; D\Big] - \max_a \mathbb{E}[V(a, \theta) \mid D] \geq 0. \tag{21.1}$$

The inequality follows because choosing after observing θ includes the option to choose the current best action.

A proposed test costing C cannot be justified solely by decision information if $C > \mathrm{EVPI}$, under the same value model. Real tests provide partial rather than perfect information, so their

expected information value is at most this upper bound. Tests may have direct product benefits, but those should be counted separately.

This prevents unlimited spending in the name of learning. When every plausible result would lead to the same decision, the decision value of further measurement is small. A new experiment is useful when it can change what the studio will do.

21.3 Sensitivity of population trajectories

Consider two deterministic models $\dot{x} = F(x, u)$ and $\dot{\tilde{x}} = F(\tilde{x}, \tilde{u})$ with

$$\|F(x,u) - F(y,v)\| \le L_x\|x - y\| + L_u\|u - v\|. \tag{21.2}$$

Gronwall's inequality gives

$$\|x(t) - \tilde{x}(t)\| \le e^{L_x t}\|x(0) - \tilde{x}(0)\| + L_u \int_0^t e^{L_x(t-s)}\|u(s) - \tilde{u}(s)\|\,\mathrm{d}s. \tag{21.3}$$

This is a sensitivity bound, not evidence that L_x is small. Near positive feedback or capacity failure, small errors can amplify quickly. Long-horizon precise forecasts then require stronger evidence than short-horizon local estimates.

21.4 A finite-horizon retained-value gradient

For acquisition A_c fixed independently of parameter θ,

$$J(\theta) = \sum_{c=0}^{T} A_c \sum_{a=0}^{T-c} \delta^{c+a} r_c(a;\theta)\, m_c(a;\theta). \tag{21.4}$$

Finite sums permit termwise differentiation:

$$J'(\theta) = \sum_{c} A_c \sum_{a} \delta^{c+a}\big[m_c(a;\theta)\,\partial_\theta r_c(a;\theta) + r_c(a;\theta)\,\partial_\theta m_c(a;\theta)\big]. \tag{21.5}$$

This separates retention and monetization effects. If acquisition also changes with θ, an additional derivative term is required. Omitting that term while changing metadata and gameplay together can misattribute the source of value.

A finite difference estimates this gradient only locally:

$$\widehat{J}'(\theta) \approx \frac{\widehat{J}(\theta + \Delta) - \widehat{J}(\theta - \Delta)}{2\Delta}. \tag{21.6}$$

Randomization supports the comparison; small Δ reduces approximation bias but can worsen signal-to-noise. Larger changes may cross qualitatively different player behaviors.

21.5 Model selection is a prediction-and-decision task

Adding parameters always allows a flexible model to fit historical data more closely. It does not necessarily improve future decisions. Use forward-chaining validation: train on earlier cohorts and evaluate on later mature cohorts. Randomly splitting rows from the same player across training and validation leaks information.

A decision-aware loss can penalize wrong actions rather than only forecast error:

$$\mathcal{L}_{\text{decision}} = J(a^*(\theta), \theta) - J(\widehat{a}, \theta). \tag{21.7}$$

A model can have slightly worse average prediction error but better decisions if it estimates the relevant boundary more accurately. Conversely, a visually excellent DAU fit can be useless for deciding whether a purchase bug should block relaunch.

21.6 Second variation of retained session value

The hazard functional admits a useful curvature calculation. Assume a finite horizon, bounded integrable v, h, η, $v \geq 0$, and perturbations small enough to keep the hazard nonnegative. Dominated convergence justifies differentiating under the integral twice:

$$\left.\frac{\mathrm{d}^2 V[h + \epsilon\eta]}{\mathrm{d}\epsilon^2}\right|_0 = \int_0^H v(t)\, S_h(t) \left(\int_0^t \eta(s)\,\mathrm{d}s\right)^2 \mathrm{d}t \geq 0. \tag{21.8}$$

Hence V is convex as a functional of hazard on its feasible convex domain. Its mixed second-variation kernel is

$$\frac{\delta^2 V}{\delta h(s)\,\delta h(r)} = \int_{\max(s,r)}^H v(t)\, S_h(t)\,\mathrm{d}t. \tag{21.9}$$

This positive-semidefinite kernel describes interaction between changes at different session times. Reducing two sequential hazards can preserve more downstream value than adding their isolated first-order effects. It does not establish that two real features have those hazard effects; that remains an empirical question.

The result also warns against assuming a globally concave development objective. Even a smooth, interpretable retention model can produce nonconcave allocation problems after controls are composed with hazard reductions and costs. Local optimization and global optimality are different claims.

21.7 An adjoint for cohort-age transport

The age-structured model can be differentiated without estimating a separate finite difference for every control. Transpose the population into a column $m(a, t)$ and consider

$$\partial_t m + \partial_a m = Q(a, t; \theta)^\top m, \qquad m(0, t) = A(t; \theta)\, e_A. \tag{21.10}$$

For a finite age interval $0 \leq a \leq a_{\max}$ and fixed initial population, define

$$J(\theta) = \int_0^T \int_0^{a_{\max}} w(a, t; \theta)^\top m(a, t)\,\mathrm{d}a\,\mathrm{d}t. \tag{21.11}$$

The vector w may weight active player-days or declared receipts. Outflow at $a_{\max}$ is allowed. Assume sufficient smoothness for the following integration by parts, or interpret the equations weakly with compatible boundary conditions.

Let the adjoint column p solve backward:

$$-\partial_t p - \partial_a p = Qp + w, \tag{21.12}$$

$$p(a, T) = 0, \qquad p(a_{\max}, t) = 0. \tag{21.13}$$

Multiplying the linearized forward equation by p and integrating by parts cancels all interior variations of m. The derivative becomes

$$\begin{aligned}\frac{\mathrm{d}J}{\mathrm{d}\theta} = &\int_0^T \int_0^{a_{\max}} \Big[p^\top (\partial_\theta Q^\top) m + (\partial_\theta w)^\top m\Big] \,\mathrm{d}a \,\mathrm{d}t \\ &+ \int_0^T p(0,t)^\top e_A \,\partial_\theta A(t;\theta) \,\mathrm{d}t.\end{aligned} \tag{21.14}$$

The first term values changed transitions; the second values changed per-state outcomes; the boundary term values changed acquisition. Initial-population or terminal-payoff dependence would add corresponding terms.

This is the continuous counterpart of the cohort sensitivity decomposition. It supplies an efficient gradient for a discretized planning model. It does not make uncertain transition rates known. Numerical implementations should compare the adjoint gradient with finite differences on a small grid before using it for allocation.

21.8 Likelihood, matrix exponentials, and identifiability

If states are actually observed at interval Δ, a homogeneous generator induces $P(\Delta) = e^{Q\Delta}$. With transition counts N_{ij}, conditional log-likelihood is

$$\ell(\theta) = \sum_{i,j} N_{ij} \log P_{ij}(\Delta;\theta). \tag{21.15}$$

For a scalar parameter, the matrix-exponential derivative is

$$\partial_\theta e^{Q\Delta} = \int_0^\Delta e^{Qs} (\partial_\theta Q)\, e^{Q(\Delta - s)} \,\mathrm{d}s. \tag{21.16}$$

Multiplying $\partial_\theta Q$ by $\Delta e^{Q\Delta}$ is valid only under an appropriate commutation condition. Matrix noncommutativity is consequential, not a notational detail.

For an origin-state exposure count N_i, the conditional expected information has form

$$\mathcal{I}(\theta) = \sum_i N_i \sum_j \frac{\nabla_\theta P_{ij} \, \nabla_\theta P_{ij}^\top}{P_{ij}}, \tag{21.17}$$

over positive-probability transitions. A rank-deficient information matrix indicates locally indistinguishable parameter directions. More observations of the same uninformative measurement do not resolve that rank deficiency.

In ordinary game telemetry, dormant and effectively lost states are not directly observed as distinct labels. The likelihood above therefore applies only when the state definitions and observations support it. Otherwise use an observation model, simpler estimands, or bounds. Fitting a detailed generator to aggregate DAU alone does not identify all its rates.

Practical implication. Use proofs to delimit what is possible, then use data to decide which assumptions are credible. Mathematical complexity is useful only when it narrows a consequential uncertainty.

Table 6. A conceptual instrumentation contract. Event names are illustrative, not exact platform API calls.

Stage	Evidence event	Failure distinction
Arrival	Session start, load complete	Load timeout or crash
Understanding	First unaided core action	Control error, help request
Loop	First reward banked	Lost reward, interrupted loop
Progression	First upgrade delivered	Insufficient funds versus failed grant
Return	New session with cohort age	New acquisition versus reactivation
Shop	Product view and prompt	Abandoned prompt versus payment failure
Purchase	Receipt reconciled, entitlement saved	Paid but not delivered

22 The event system and an end-to-end experimental protocol

The most valuable analytics work often happens before the first advertisement. A launch should produce a reliable sequence from assignment to exposure, action, reward, progression, return, and purchase delivery. Without that sequence, the studio may acquire thousands of users yet remain unable to identify the largest obstacle.

22.1 An event contract

A conceptual event record is

$$e = (i, s, t_e, t_{\text{ingest}}, v, z, k, a, o, \nu), \tag{22.1}$$

where i is a pseudonymous player key, s session, v build/config version, z source or experiment stratum, k event kind, a relevant action identifier, o outcome, and ν schema version. This is an analytical schema, not a requirement that every field be sent to Roblox custom analytics. Platform field and retention limits must be respected.

Collect only what the experiment needs. Aggregate behavior is usually sufficient for design decisions; private messages, unnecessary personal details, or external identity data are not required for this blueprint.

22.2 Funnel and economy consistency

Roblox provides funnel and economy event systems [16, 17]. A first-loop funnel should reflect successful stages. A reward granted event belongs after the authoritative operation succeeds, with an idempotent operation identity where the system needs deduplication.

Maintain invariants:

$$\#\text{delivered unique receipts} \le \#\text{recognized paid receipts}, \tag{22.2}$$

$$\text{ending wallet} = \text{starting wallet} + \text{sources} - \text{sinks}, \tag{22.3}$$

$$\text{cohort first arrival} = \min(\text{eligible historical arrivals}). \tag{22.4}$$

Algorithm 1 Release decision procedure

```
Freeze cohort and experiment definitions
Verify assignment and event reconciliation on test traffic
Enroll eligible units; persist assignment
Observe actual build exposure and operational guardrails
if a declared critical failure occurs then
    Stop immediately
else
    Collect the prespecified sample and follow-up
end if
Estimate the primary contrast with design-aware uncertainty
Inspect declared device/source heterogeneity
Choose rollout, revision, or further targeted measurement
Archive the decision and the exact evidence window
```

The first inequality is a diagnostic conceptual constraint; refunds, external grants, and migration procedures need separately labelled records. A raw event count can exceed receipt count if logging is retried, which is why unique operation identity matters.

22.3 The experiment specification

A complete specification states:

1. The population, enrollment trigger, assignment unit, and persistence rule.
2. The product difference, including all affected configuration values.
3. One primary estimand and the minimum useful effect.
4. Follow-up duration, sample plan, and valid stopping rule.
5. Guardrails and rollback conditions.
6. Handling of missing events, old server versions, and repeat users.
7. The decision to be taken under favorable, neutral, harmful, or inconclusive evidence.

This is a compact scientific protocol, not paperwork for its own sake. It prevents changing the definition of winning after seeing the result.

22.4 Pseudo-algorithm for a release decision

The algorithm intentionally separates operational safety from statistical superiority. A critical entitlement failure should not continue simply because D1 requires another day of follow-up.

22.5 End-to-end extraction

A final launch extraction should include campaign identity, delivery dates, spend, impressions, clicks, qualified acquisition, early-session behavior, exact-day cohorts, player-days, payer conversion, receipts, reward streams, source composition, devices, and errors. Use a window covering the original launch and its complete follow-up.

Keep the extraction reproducible: retain source links, capture times, table definitions, and transformation logic. A chart generated from several views should identify every join key and unit conversion. A raw archive without a readable interpretation is incomplete; a polished narrative without recoverable evidence is equally incomplete.

Practical implication. Instrument the questions the studio must answer after spending. The goal is not to collect everything; it is to make the next development and acquisition decisions identifiable.

23 An illustrative new-game design: the orchard escape loop

Consider a hypothetical game in which players chop trees, collect seeds, escape pursuit, plant at a base, and use earnings to improve tools. This example is a design exercise, not a claim of market novelty or a forecast of profitability. It translates the framework into concrete choices.

The smallest complete loop is

$$\text{choose tree} \to \text{chop} \to \text{carry seed} \to \text{escape} \to \text{plant} \to \text{choose next upgrade}. \tag{23.1}$$

Each arrow should be observable and understandable. The first session should reach one complete loop quickly enough to reveal why the next loop could be different.

23.1 A state model for meaningful decisions

Let state be $x = (a, s, v, d, \tau, g)$: axe capability a, movement skill or speed s, carried value v, distance to safety d, pursuit state τ, and planted-garden state g. Actions include chop again, bank now, choose another route, use an earned ability, or seek help.

A finite-horizon decision model is

$$V(x) = \max_{u \in U(x)} \Big\{ r(x,u) + \delta \sum_{x'} P(x' \mid x, u)\, V(x') \Big\}. \tag{23.2}$$

The game does not need to solve this equation during play. The designer uses it to ask whether multiple actions can be rational in different states. If one action dominates everywhere, the apparent choice is not meaningful.

23.2 Traits instead of only rarity

A seed may encode growth time, shape, yield profile, environmental affinity, or an interaction with another planted tree. Let trait vector be $\theta \in \mathbb{R}^k$. Garden value could depend on combinations:

$$Y(g) = \sum_{i} y(\theta_i) + \sum_{i<j} \theta_i^{\top} W \theta_j. \tag{23.3}$$

The interaction matrix W is a design object. Positive and negative interactions can create placement choices, but must remain readable. Hidden arbitrary synergies do not create fair strategic depth.

Rarity should not be the only determinant of usefulness. A common tree with a distinct role can remain relevant after a rare acquisition. This preserves horizontal choice and reduces the pressure to replace every old item.

23.3 The axe and speed economy

If chop time is $t_{\text{chop}}(a, \text{tree})$ and escape probability is $p_{\text{escape}}(s, d, \tau)$, expected banked value per attempted run is

$$\mathbb{E}[V_{\text{bank}}] = p_{\text{escape}}\, v \tag{23.4}$$

under a simple all-or-nothing model. Expected value per unit time divides by chop, carry, travel, and reset time. Faster axes can increase risk exposure if they encourage deeper excursions; the interaction should be tested rather than assumed beneficial.

An upgrade should change an action or reachable option, not merely shorten a wait. Examples include choosing between a heavy axe with a stronger final chop and a lighter axe allowing safer movement, or a garden trait changing which seeds are desirable.

23.4 A first-session test

Compare an opening that immediately presents a small readable tree and a visible safe return path with an opening requiring several separate training steps first. Measure load completion, first chop, first bank, first planting, and second voluntary run. D1 remains important, but early stage data explains where a difference emerged.

The test should preserve the actual advertised mechanic. Showing a chase in the thumbnail while making the player train in an empty room first is a plausible mismatch hypothesis. It must be tested, not asserted as the cause of every bounce.

23.5 An expansion path

New content can add a pursuit behavior, a route interaction, a seed trait, or a cooperative banking opportunity. Each should interact with the existing loop. A second currency with no distinctive choice adds complexity without necessarily adding depth.

The first release needs a reliable save, a clear first reward, a few differentiated choices, and enough variation to reveal repeat-play potential. Broader content should follow evidence about what players understood and voluntarily repeated.

Practical implication. Use the loop's state variables to design testable decisions. A promising concept becomes a product when players can understand, execute, repeat, and remember its core interaction.

24 Release engineering and the bridge from a patch to an experiment

A published patch is not a single synchronized treatment. Old servers may remain alive, players may carry state from prior builds, and configurations may change independently of code. Roblox documents that players are not automatically removed from old servers when an update is published [26]. Version exposure must therefore be measured.

24.1 Exposure-aware analysis

Let V_{it} be the actual version experienced by player i at time t. A release timestamp alone supplies an intended treatment Z_t, not proof of exposure. The contrast

$$\mathbb{E}[Y \mid t > t_{\text{release}}] - \mathbb{E}[Y \mid t < t_{\text{release}}] \tag{24.1}$$

mixes versions, audience changes, and calendar effects. Log build and configuration identity at consequential events, not just at session start if a live configuration can change mid-session.

Live configuration systems can change values without a full restart [27]. Feature flags should have an exposure rule. If the flag changes an economy, decide whether a player's assignment persists across sessions. Mixing reward rules during a single progression path can create fairness and analytical problems.

24.2 Migration as a correctness contract

Let M_v map saved state from schema v to $v+1$. A migration should preserve declared invariants:

$$I_v(x) \Rightarrow I_{v+1}(M_v(x)). \tag{24.2}$$

Idempotency is important when retries occur:

$$M_v(M_v(x)) = M_v(x) \tag{24.3}$$

after accounting for a schema marker or an equivalent guarded implementation. Applying a currency grant twice violates the intended contract even if the migrated state otherwise loads.

Backward compatibility matters while old and new servers coexist. A new schema unreadable by old servers can turn ordinary rejoining into state loss or failure. The release plan must either maintain compatibility or coordinate server transition deliberately.

24.3 Canary releases and rollback

A canary exposes a limited, declared portion of traffic while monitoring correctness and device performance. It is not necessarily powered to detect D7 retention. Its purpose is to catch high-impact failures before broad exposure.

Rollback must address both code and state. Reverting a script does not undo already granted items, changed prices, or saved progression. A rollback plan should distinguish reversible configuration, compensating actions, and irreversible state transitions.

A release is ready only if the team can answer what happens when a purchase arrives during migration, a server closes during a save, or an old client requests an action against new rules. These are concrete product paths, not hypothetical bureaucracy.

24.4 Separating novelty from improvement

An update can improve technical reliability while its novelty simultaneously increases visits. To estimate the reliability effect, compare actual operation success and matched cohort outcomes rather than only total activity. To estimate the event effect, track returning and new users separately.

A release log should include hypothesis, version, time, affected systems, known limitations, expected metric direction, and rollback trigger. This creates a historical dataset for estimating production and update effects rather than relying on memory.

24.5 A cadence the team can sustain

A weekly promise that repeatedly ships broken changes can be worse than a slower reliable rhythm. Let h be available hours and c_{fix} emergency work. Planned content capacity is $h - c_{\text{fix}}$. Reducing incident work can create more update capacity than hiring another content contributor.

The bridge to live operations is therefore a reliable production system. Self-renewing content still requires monitoring, curation, and occasional structural improvement. It does not mean unattended code or unlimited algorithmic content generation.

Practical implication. Treat every patch as both a product intervention and a versioned exposure event. Reliable deployment makes its player effect measurable and its success repeatable.

25 Novelty, memory, and the production of new situations

An update can supply more assets without supplying more meaningful play. The relevant resource is the set of situations a player can still discover, interpret, and influence. That set depends on rules, maps, opponents, items, goals, and the player's own accumulated knowledge.

Let an encounter be generated from components $C_1, \dots, C_k$. The theoretical combination count is

$$|\Omega| = \prod_{j=1}^{k} |C_j|. \tag{25.1}$$

This count is an upper bound on variety, not experienced depth. Invalid combinations, dominant strategies, unreadable outcomes, and repeated matchmaking can collapse the useful set. Let $\Omega_{\text{viable}} \subseteq \Omega$ and encounter probabilities p_ω. Effective diversity is

$$N_{\text{eff}} = \exp\left(- \sum_{\omega \in \Omega_{\text{viable}}} p_\omega \log p_\omega \right). \tag{25.2}$$

Adding a component matters when it changes the experienced distribution and the decisions available within it.

25.1 Novelty is player specific

A new player can find an old map novel. A veteran may exhaust a new item in one session. Define $N_i(t)$ as a model of player i's remaining meaningful novelty, not a shared currency consumed by the entire server. A simple between-update approximation is

$$\dot{N}_i = -d_i N_i, \qquad N_i(t_k^+) = N_i(t_k^-) + A_{ik}. \tag{25.3}$$

The impulse A_{ik} depends on whether the update creates unfamiliar and useful situations for that player. A cosmetic recolor and a new route interaction can have very different A_{ik} even if they take equal production time.

A more expressive model incorporates learning and recombination:

$$\dot{N}_i = s_i(g_t, \text{social state}) - d_i N_i - \ell_i(N_i, \text{play rate}), \tag{25.4}$$

where s_i is endogenous creation of meaningful situations by the system. This is the opening for updates that continue working after release: they can raise s_i by expanding how existing components interact.

25.2 New content can amplify old content

Suppose a new map m supports r distinct role interactions and v viable movement or tool choices. Its marginal interaction opportunities may scale with rv, rather than one asset. But multiplying counts is insufficient when those combinations feel equivalent.

A measurable interaction lift compares a new component both alone and with the existing system. For outcome Y, a factorial contrast is

$$\tau_{\text{interaction}} = \mathbb{E}[Y_{11} - Y_{10} - Y_{01} + Y_{00}]. \tag{25.5}$$

A positive interaction means the combination creates more value than the sum of isolated changes under the test. This is a stronger claim than saying that many combinations are theoretically possible.

25.3 Memory and recognizable identity

Updates also draw value from familiarity. Players know controls, understand goals, recognize social roles, and remember prior achievements. A complete redesign can increase novelty while destroying this accumulated competence.

Let M_i represent useful familiarity. A stylized experience value is

$$U_i = \alpha \log(1 + N_i) + \beta \log(1 + M_i) - \gamma \,\text{confusion}_i. \tag{25.6}$$

The coefficients are preference assumptions. The expression captures a design tension: a game can be too repetitive or too disorienting. A successful update may preserve the recognizable core while changing the situations encountered within it.

This suggests a modular architecture. Keep input grammar, role clarity, persistence, and reward meaning stable where possible. Add content through maps, modifiers, collections, and optional goals that the existing grammar can express. When the grammar itself needs repair, provide migration and re-teaching rather than disguising a redesign as ordinary content.

25.4 Social stories as renewable variation

Other players create uncertainty that authored content does not fully enumerate: a mistaken suspicion, an unexpected escape, a coordinated rescue, or an unusual trade. These events can refresh a stable mechanical loop. They require legible rules and fair opportunities; chaos alone is not a retention system.

Measure whether a release increases varied viable choices, intentional co-play, repeat rounds, and voluntary return. A rise in item acquisition without these changes may indicate a collection pulse rather than a deeper loop.

Practical implication. Design updates that add interactions to familiar systems. A component earns recurring value when players continue making new decisions with it after its initial announcement is forgotten.

26 Update cadence and stability of the retention system

Cadence should follow the decay of player value, production capacity, and the effect of updates on future transitions. A fixed weekly rhythm is not inherently optimal. Neither is an infrequent large expansion. The relevant comparison includes content quality, fatigue, reliability, and the community's ability to return together.

26.1 An impulse-decay model

Suppose novelty is reset to A at each update and decays as Ae^{-dt} between updates spaced τ apart. Average novelty is

$$\bar{N}(\tau) = \frac{A(1 - e^{-d\tau})}{d\tau}. \tag{26.1}$$

This is a reset model. An additive-impulse model has a different steady-state average; the assumptions must not be mixed.

Let a release cost C, novelty carry weight w, and other cadence costs be temporarily omitted:

$$U(\tau) = \frac{wA(1 - e^{-d\tau})}{d\tau} - \frac{C}{\tau}. \tag{26.2}$$

Then

$$U'(\tau) = \frac{1}{\tau^2}\left[C - \frac{wA}{d}\big\{1 - (1 + d\tau)e^{-d\tau}\big\}\right]. \tag{26.3}$$

An interior stationary cadence satisfies

$$C = \frac{wA}{d}\big\{1 - (1 + d\tau)e^{-d\tau}\big\}. \tag{26.4}$$

For $0 < C < wA/d$, the right side increases from zero to wA/d, giving one stationary maximum. If release quality A depends on preparation time, or bugs increase with frequency, the equation changes. It is a sensitivity model, not a universal release calendar.

26.2 Retention stability across release cycles

Let $x_k = (A_k, D_k)^\top$ be active and dormant populations just after update k. Within-cycle transitions yield a nonnegative substochastic operator P_τ. At the next update, reactivation is represented by

$$R_\rho = \begin{pmatrix} 1 & \rho \\ 0 & 1 - \rho \end{pmatrix}. \tag{26.5}$$

For this column-vector convention, column sums are one. Pure reactivation conserves mass; it cannot create new players.

If social activity additionally creates new eligible acquisitions, represent the expected incremental arrivals by nonnegative operator K_τ. Then an illustrative cycle map is

$$x_{k+1} = B_\tau x_k + a_k, \qquad B_\tau = R_\rho P_\tau + K_\tau. \tag{26.6}$$

K_τ must count genuinely new incremental players, not existing friends returning under a different label. External acquisition is a_k.

Proposition 26.1 (Subcritical cycle stability)**.** *If B_τ is nonnegative, has spectral radius $\varrho(B_\tau) < 1$, and $a_k = a$, the unique fixed point is*

$$x^* = (I - B_\tau)^{-1}a, \tag{26.7}$$

and $x_k \to x^*$ from every finite initial state.

Proof. The spectral-radius condition implies convergence of the Neumann series $\sum_{j\geq 0} B_\tau^j = (I - B_\tau)^{-1}$. Iteration gives $x_k = B_\tau^k x_0 + \sum_{j=0}^{k-1} B_\tau^j a$. The first term vanishes and the second converges to the stated fixed point. □

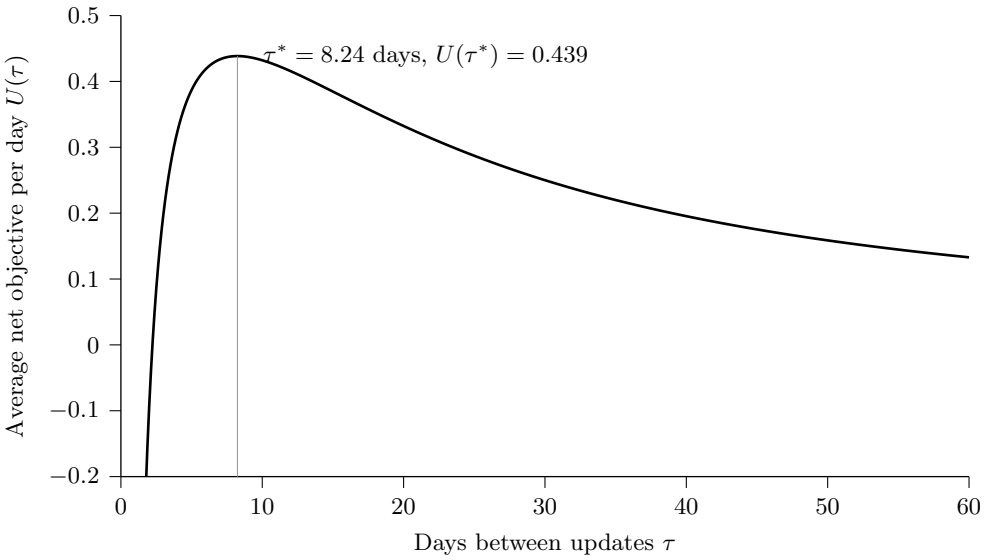


Figure 10. Synthetic cadence objective $U(\tau)$ of equation (26.2) with $A = 1$, $d = 0.1$ per day, $w = 1$, and release cost $C = 2$ in compatible objective units; the thin vertical rule marks the stationary point of equation (26.4). The preferred interval is conditional on this simplified reset model. Drawn with PGFPLOTS [32]; stationary point found with SciPy [34].

Corollary 26.2 (Reactivation alone is not unbounded growth)**.** *If $K_\tau = 0$ and P_τ is column substochastic, then $\varrho(R_\rho P_\tau) \le 1$.*

Proof. The induced matrix 1-norm of a nonnegative column-substochastic matrix is at most one. R_ρ has norm one. Submultiplicativity and $\varrho(B) \le \|B\|_1$ yield the claim. □

An update can improve persistence and recover dormant players without generating unlimited growth. This matters when a studio interprets a large return pulse as evidence of self-sustaining acquisition.

If $\varrho(B_\tau) > 1$, the linear model predicts growth, but finite audiences and server capacity eventually invalidate linearity. The spectral radius is a local diagnostic under the model, not a published Roblox ranking threshold.

For a synthetic stability map, take

$$P_\tau = \begin{pmatrix} 0.80 & 0.05 \\ 0.10 & 0.85 \end{pmatrix}, \qquad K_\tau = \begin{pmatrix} k & 0 \\ 0 & 0 \end{pmatrix}, \tag{26.8}$$

using the column-vector convention. Vary $0 \le \rho \le 0.8$ and $0 \le k \le 0.35$. The resulting boundary depends on these assumed entries, not on a measured Roblox threshold.

26.3 Which improvement changes stability most?

For a simple dominant eigenvalue with right vector v and left vector w, normalized by $w^\top v = 1$,

$$\frac{\partial \varrho(B)}{\partial \theta} = w^\top \frac{\partial B}{\partial \theta} v. \tag{26.9}$$

This sensitivity identifies influential transitions within the fitted local model. A large derivative is meaningful only if the corresponding parameter can actually be changed and estimated.

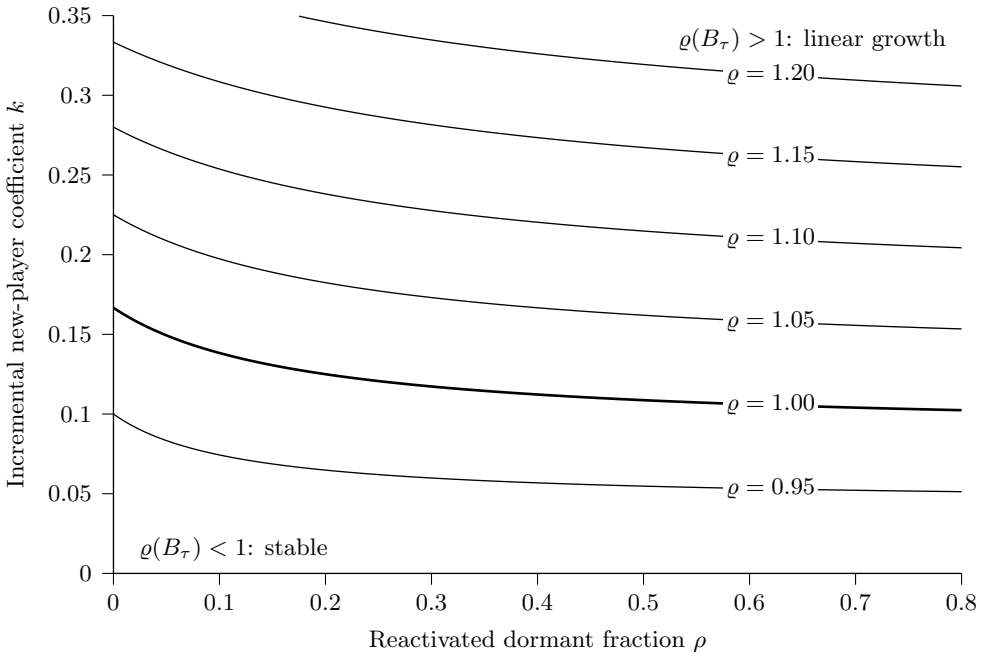


Figure 11. Isolines of the dominant cycle eigenvalue $\varrho(B_\tau)$ for the operators of equation (26.8), computed in closed form from the characteristic polynomial of B_τ. The heavy line is $\varrho = 1$, the linear model's stability boundary; thin lines are spaced 0.05 apart. With $k = 0$ the dominant eigenvalue equals 0.90 for every ρ, so the $\varrho = 0.90$ isoline coincides with the horizontal axis. Finite-audience and capacity constraints are omitted. Drawn with PGFPLOTS [32]; isolines computed in closed form and checked with NumPy [33].

Practical implication. Choose cadence by how long updates create useful play, how safely the team can ship, and which transitions improve. A return pulse, a healthier kernel, and incremental acquisition are three different accomplishments.

27 Murder Mystery 2 as a structural retention case

Murder Mystery 2 provides a useful design reference because a recognizable round structure coexists with collection and trading. Its official public description identifies Innocent, Sheriff, and Murderer roles and a knife collection/trading layer [29]. The following analysis is structural inference, not access to its private retention, revenue, update schedule, or experimental results.

No numerical D1, D7, payer-conversion, or causal update effect is asserted for Murder Mystery 2. Public longevity and familiar mechanics do not reveal which individual feature caused its success. Survivorship bias also matters: copying visible features of a surviving game does not reproduce its history, audience, or execution.

27.1 A stable grammar, variable situations

The round supplies a compact grammar: roles have distinct objectives, other players hold uncertain information, and a new round creates another chance to interpret and act. The source of variation is not only a new map. Different players, movement, suspicion, timing, and role assignments can change the same map's situation.

Let a round state be

$$s = (m, r, \ell, p, h), \tag{27.1}$$

where m is map, r role configuration, ℓ spatial configuration, p player strategies, and h information history. A game can create many experienced situations even when the set of authored maps is modest. The important quantity is the distribution of meaningful decisions over these states.

A new map potentially changes sight lines, encounter timing, escape options, and information flow. It can therefore interact with several existing roles. Whether it actually improves retention remains a hypothesis requiring player evidence.

27.2 Short loops and long goals

A round gives a short opportunity for action and a clear reset. A collection or mastery goal can persist across rounds. These operate at different timescales:

$$V_i = V_i^{\text{round}} + V_i^{\text{mastery}} + V_i^{\text{collection}} + V_i^{\text{social}}. \tag{27.2}$$

This is a conceptual decomposition, not an additive empirical fit. The terms can interact. A cosmetic collection may create identity in a social setting, while a memorable round may make the associated item more meaningful.

The short loop must remain worthwhile without completing the long goal. If all enjoyable play is deferred until a rare acquisition, the collection layer becomes a gate rather than a complement. A durable game gives new and returning players meaningful action at their current progression state.

27.3 Updates that reuse the grammar

A hypothetical seasonal update to this type of game can add a map, a themed collection, optional objectives, or a modifier. Each addition can reuse the existing role grammar. Players do not need to relearn the entire product to understand why the new content matters.

Consider a content component c with implementation cost K_c and interaction set $\mathcal{I}(c)$. A useful design efficiency measure is

$$E_c = \frac{\sum_{s \in \mathcal{I}(c)} \Pr(s \text{ experienced}) \, \Delta V(s)}{K_c + C_c^{\text{maintenance}}}. \tag{27.3}$$

This is a proposed planning measure. It values the quality and likelihood of interactions, not only their count. A map that creates many broken or unreadable encounters has low or negative ΔV.

27.4 Collection without destroying the round

An item layer can provide expression, memory, and goals while leaving the round's essential information and counterplay understandable. Changes that make item ownership dominate outcomes risk weakening the very social encounters that give the collection a stage.

A balancing test should compare participation and outcomes across progression states. If new users consistently encounter an unwinnable or unreadable environment, long-term collectors may be retained while the acquisition funnel deteriorates. Aggregate veteran activity can conceal that failure.

Trading also creates interaction and identity, but introduces duplication, delivery, eligibility, and trust responsibilities. A reliable item ledger and transparent restrictions are part of the product, not a separate analytics concern.

27.5 What transfers and what does not

The transferable pattern is a stable understandable loop with new combinations, persistent identity, and social variation. The nontransferable assumption is that adding knives, rarity, or a seasonal banner will reproduce the same audience response.

An orchard escape game can use the pattern without reproducing the theme: preserve chopping, carrying, escaping, and planting as the grammar; vary pursuit behavior, route interactions, seed traits, and garden choices. The question is whether each update creates useful decisions within that grammar.

Practical implication. Study the relationship between core loop and update layer. Borrow a mechanism to test, not a claim that another game's visible features guarantee retention.

28 Designing updates that continue working after release

The final operational objective is an update whose value survives its announcement. The release should do more than temporarily increase arrivals. It should improve the system that produces future sessions: clearer choices, more viable combinations, reliable progression, richer social situations, and reasons to return that remain understandable.

28.1 A three-stage update contract

Separate reach, reactivation, and persistence. Let R_k be returning players reached by update k, q_k the fraction who return, and $p_k(a)$ their subsequent participation kernel. Incremental post-update player-days are modeled as

$$\Delta P_k(H) = R_k q_k \sum_{a=0}^{H} p_k(a) - P_k^{\text{counterfactual}}(H). \tag{28.1}$$

The counterfactual subtracts activity that would have occurred anyway. Without a credible comparison, report gross post-update activity rather than claiming incrementality.

An update can win reach while failing persistence. It can also improve persistence without a dramatic spike. Evaluate both. Opted-in event and update announcements are distribution tools described by Roblox [25]; they are not substitutes for a worthwhile returning session.

28.2 A reusable update architecture

Maintain a stable core with explicit extension points:

1. **Encounter extension.** Add a map feature, route, role interaction, or pursuit behavior that changes a readable decision.

2. **Progression extension.** Add a horizontal goal or a new use for existing earned progress without invalidating prior effort.

3. **Social extension.** Create a cooperative, competitive, or expressive interaction worth sharing and repeating.

4. **Return extension.** Give dormant players a clear reason to re-enter and a short path to experiencing the change.

5. **Reliability extension.** Improve the systems supporting every future loop: loading, persistence, matchmaking, and entitlement delivery.

A release need not contain all five. It should identify which one it actually improves and how that improvement will be measured.

28.3 An update hypothesis with a measurable mechanism

For a hypothetical orchard game, suppose a new pursuit behavior forces a choice between a shorter exposed route and a longer sheltered route. A seed trait interacts with that choice, while a garden objective rewards banking either of two useful seed types. The update extends existing systems rather than adding a disconnected minigame.

The hypothesis is not merely that DAU rises. It is that route choice becomes less dominated, more users voluntarily complete a second run, and returning players continue using the new interaction after the initial event. Measure:

$$\text{choice diversity} = -\sum_{a} p_a \log p_a, \tag{28.2}$$

$$\text{second-loop conversion} = \frac{\#\text{players completing a second voluntary loop}}{\#\text{players completing the first}}, \tag{28.3}$$

$$\text{post-update player-days} = \sum_{d=0}^{H} r_{\text{update}}(d). \tag{28.4}$$

Choice diversity alone is not a success measure: random confusion can also increase entropy. Pair it with successful execution, player understanding, and return outcomes.

28.4 A post-update evidence sequence

Immediately verify delivery, migrations, error rates, and performance. After processing delay, inspect exposure and first-session outcomes. When cohorts mature, compare exact-day returns and player-days. Examine whether collection, shop, and social systems support the improved loop without displacing it.

Compare new players, returning active players, and reactivated dormant players separately. A change can help veterans and confuse newcomers. Preserve version identity so that the observed effect is not diluted by old servers or partial rollout.

If the update underperforms, classify the failure: insufficient reach, unclear promise, failed first experience, shallow content, poor persistence, technical regression, or incompatible economy. The next action should target that mechanism. Repeating a larger announcement does not repair an unidentified product failure.

A synthetic daily example makes the pulse-versus-persistence distinction explicit. Use row-vector transitions

$$P_0 = \begin{pmatrix} 0.55 & 0.35 \\ 0.02 & 0.90 \end{pmatrix}, \qquad P_1 = \begin{pmatrix} 0.70 & 0.25 \\ 0.04 & 0.93 \end{pmatrix}. \tag{28.5}$$

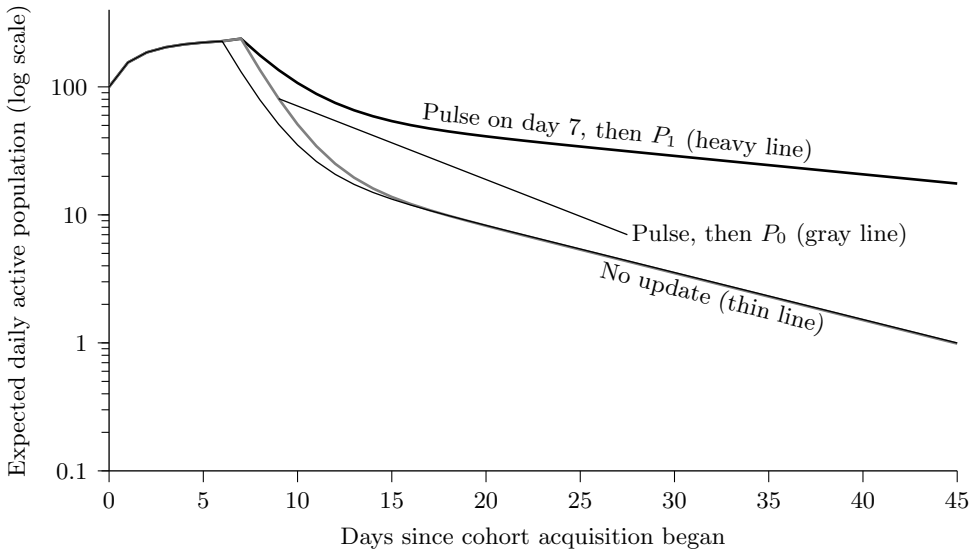


Figure 12. Synthetic update trajectories on a logarithmic vertical axis: no update (thin black line), reactivation pulse on day 7 followed by P_0 (gray line), and the same pulse followed by P_1 (heavy black line). After acquisition stops, each tail decays geometrically at the dominant eigenvalue of its transition matrix (0.919 for P_0, 0.967 for P_1) and therefore appears as a straight line. The pulse alone rejoins the no-update tail by about day 20, because players moved into the active state leave it faster under P_0 than they would have left the dormant state; only the persistent improvement changes the slope. All model parameters are stated above. Drawn with PGFPLOTS [32]; trajectories computed with NumPy [33].

Missing row mass exits the modeled population. Acquire 100 players daily on days 0–6, then stop acquisition. On day 7 reactivate 30 percent of dormant players. Compare that pulse alone with the pulse followed by P_1 replacing P_0. These are deliberately specified scenarios, not fitted effects.

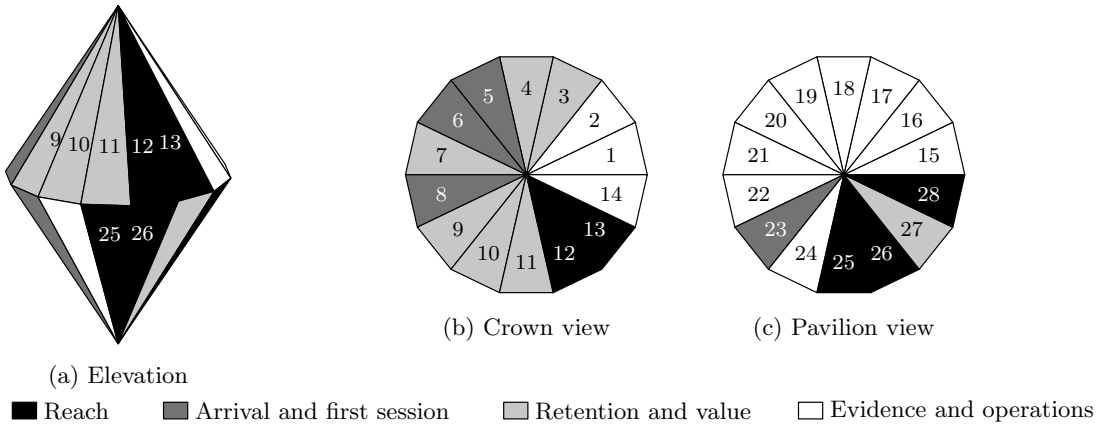
28.5 The retention engine

A repeatable release system can be expressed as

$$\begin{aligned} \text{friction} &\to \text{design change} \to \text{reliable release} \\ &\to \text{meaningful situations} \to \text{measured return} \\ &\to \text{next design choice.} \end{aligned} \tag{28.6}$$

This is a feedback policy with a stable product identity. Developers supply and curate the rules, assets, goals, and safeguards. Players generate combinations, mastery, relationships, and stories within them. The next update adds to that possibility space while preserving what made the previous sessions legible.

The work after release is specific: inspect the transition the update was meant to improve, preserve the combinations players actually use, remove broken or dominated options, and build the next addition around the remaining friction. When the familiar loop remains satisfying and the new component keeps changing what players can do together, the update continues producing reasons to return long after its launch-day traffic has passed.



(a) Elevation (b) Crown view (c) Pavilion view

Figure 13. The crystal model. (a) Orthographic elevation of the 14-gonal bipyramid, viewed 16° above the girdle plane; hidden facets are omitted and only the wider visible facets are numbered. (b) Crown view from above, facets 1–14. (c) Pavilion view, drawn in the same orientation as (b), as if seen through the crown, so that pavilion facet $k+14$ lies directly beneath crown facet k and the two share girdle edge k. Fill shades give the funnel group of Table 7: black, reach; dark gray, arrival and first session; light gray, retention and value; white, evidence and operations. Drawn by the author with TikZ [31]; projection and back-face removal computed with NumPy [33].

29 The crystal model: a 28-facet synthesis

The preceding 28 sections each isolate one structure of a live game: a measurement, a transition, a constraint, or a design lever. Read separately, they can look like a checklist. This section introduces the *crystal model*, an organizing construct developed by the author, in which the 28 structures are the 28 facets of a single crystal. The construct has one purpose: to show that click-through, the most visible acquisition number, is only the light entering the crystal. What that light becomes depends on every facet it passes through.

The crystal model is a synthesis device, not an empirically validated law. Its multiplicative form is an assumption stated in Section 29.3, and Section 29.8 gives the observation that would falsify it for a particular game.

29.1 Geometry of the crystal

Take the 14-gonal bipyramid: a regular 14-gon (the *girdle*) with one apex above and one below. It has

$$V = 14 + 2 = 16, \qquad E = 14 + 2 \cdot 14 = 42, \qquad F = 2 \cdot 14 = 28, \tag{29.1}$$

and satisfies Euler's relation $V - E + F = 16 - 42 + 28 = 2$. Every facet is a congruent isosceles triangle. The upper *crown* carries facets 1–14 and the lower *pavilion* carries facets 15–28, so facet j corresponds to Section j of this manuscript.

Crown facet k and pavilion facet $k+14$ share girdle edge k. The pairing places each structure of Sections 1–14 against one of Sections 15–28: the controlled trajectory (facet 1) meets statistical inference (facet 15); first-session survival (facet 5) meets synthetic tests (facet 19); technical reliability (facet 8) meets the event system (facet 22); creative selection (facet 12) meets update cadence (facet 26). The pairing is not a claimed causal link; Section 29.7 gives the adjacency a precise statistical role and a test of whether it earns it.

29.2 Facet scores and the four facet groups

Each facet j receives a score $s_j \in (0, 1]$ computed from a declared proxy under a measurement contract of the form (2.6). A score of 1 means the structure is, for the studio's declared purpose, as good as it needs to be; a score near 0 means it is failing. Table 7 lists the facets, their sections, an example proxy for each, and the funnel group to which each is assigned:

- $\mathcal{R}$ *Reach* (5 facets): structures that move the impression-to-click stage.
- $\mathcal{A}$ *Arrival and first session* (4 facets): structures that move a click to a qualified, successful first loop.
- $\mathcal{V}$ *Retention and value* (7 facets): structures that move a first loop to repeat play and delivered value.
- $\mathcal{E}$ *Evidence and operations* (12 facets): structures that determine how accurately the other scores are known and how safely changes can be made.

The groups partition the facets: $|\mathcal{R}| + |\mathcal{A}| + |\mathcal{V}| + |\mathcal{E}| = 5 + 4 + 7 + 12 = 28$. The assignment records each facet's primary channel; several structures plainly act on more than one stage, and a studio may reassign them provided the partition is declared before scores are examined.

For a group G with weights $w_j^G > 0$, $\sum_{j \in G} w_j^G = 1$, define the group clarity and the overall clarity as weighted geometric means:

$$\Phi_G(s) = \prod_{j \in G} s_j^{\,w_j^G}, \tag{29.2}$$

$$\Phi(s) = \prod_{j=1}^{28} s_j^{\,w_j}, \qquad w_j > 0, \quad \sum_{j=1}^{28} w_j = 1. \tag{29.3}$$

The weights are studio choices. Absent a defensible reason to differ, equal weights $w_j = 1/28$ and $w_j^G = 1/|G|$ are the neutral default.

29.3 From 28 structures to retained click-through

Equation (12.2) factorizes retained users per impression into a click stage, an arrival stage, and a retention stage. Call this ratio the *retained click-through rate*,

$$\mathrm{rCTR} = \frac{R}{I} = c_{\mathrm{click}}\, q_{\mathrm{arrival}}\, q_{\mathrm{retain}}. \tag{29.4}$$

The crystal model posits that each stage equals a ceiling times the clarity of its group:

$$c_{\mathrm{click}} = \bar{c}\,\Phi_{\mathcal{R}}(s), \qquad q_{\mathrm{arrival}} = \bar{a}\,\Phi_{\mathcal{A}}(s), \qquad q_{\mathrm{retain}} = \bar{r}\,\Phi_{\mathcal{V}}(s), \tag{29.5}$$

where $\bar{c}, \bar{a}, \bar{r} \in (0, 1]$ are the stage values attainable when every facet in the group scores 1. Substituting,

$$\mathrm{rCTR} = \bar{c}\,\bar{a}\,\bar{r}\;\Phi_{\mathcal{R}}(s)\,\Phi_{\mathcal{A}}(s)\,\Phi_{\mathcal{V}}(s). \tag{29.6}$$

The twelve evidence facets do not appear in (29.6) as multipliers. They enter through the uncertainty of the estimate (Proposition 29.3). In this sense all 28 structures bear on click-through: sixteen determine what a click becomes, and twelve determine how well the studio knows it.

Table 7. The 28 facets of the crystal. Group: $\mathcal{R}$ reach, $\mathcal{A}$ arrival and first session, $\mathcal{V}$ retention and value, $\mathcal{E}$ evidence and operations. Proxies are examples drawn from the cited sections, not prescriptions.

Facet	Structure (Section)	Group	Example proxy
1	Controlled trajectory (1)	$\mathcal{E}$	Objective, horizon and constraints declared
2	Measurement contracts (2)	$\mathcal{E}$	Share of reported metrics with a full contract
3	Cohort renewal (3)	$\mathcal{V}$	Player-days per acquisition L_{30}
4	Multistate return (4)	$\mathcal{V}$	Persistence after reactivation
5	First-session survival (5)	$\mathcal{A}$	$S(5)$ or RMST(H)
6	Onboarding (6)	$\mathcal{A}$	First-loop completion within a time limit
7	Challenge and progression (7)	$\mathcal{V}$	Effective state diversity N_{eff}
8	Technical reliability (8)	$\mathcal{A}$	Successful first-loop probability q_{loop}
9	Social play (9)	$\mathcal{V}$	Repeat co-play rate
10	Economy integrity (10)	$\mathcal{V}$	Ledger residual ϵ_{ledger} near zero
11	Monetization (11)	$\mathcal{V}$	Delivered-entitlement rate and payer value
12	Creative selection and CTR (12)	$\mathcal{R}$	Qualified click-through rate
13	Discovery response (13)	$\mathcal{R}$	Home-acquired cohort response
14	Causal experiments (14)	$\mathcal{E}$	Share of decisions backed by randomization
15	Statistical inference (15)	$\mathcal{E}$	Interval width at decision time
16	Outcome maturity (16)	$\mathcal{E}$	Share of cohorts with mature follow-up
17	Production capacity (17)	$\mathcal{E}$	Planned capacity remaining after incidents
18	Constrained control (18)	$\mathcal{E}$	Hard constraints satisfied
19	Synthetic tests (19)	$\mathcal{E}$	Accounting invariants passed
20	Launch audit (20)	$\mathcal{E}$	Reconciled launch metrics
21	Analytical limits (21)	$\mathcal{E}$	Identified parameter directions
22	Event system (22)	$\mathcal{E}$	Event-to-receipt reconciliation rate
23	Core-loop design (23)	$\mathcal{A}$	Second voluntary loop conversion
24	Release engineering (24)	$\mathcal{E}$	Version-tagged exposure coverage
25	Novelty and memory (25)	$\mathcal{R}$	Returning use of new interactions
26	Update cadence (26)	$\mathcal{R}$	Returning players reached per update
27	Structural reference (27)	$\mathcal{V}$	Reuse of the core grammar by updates
28	Self-renewing updates (28)	$\mathcal{R}$	Post-update returning player-days

Equation (29.6) restates two warnings from earlier sections in one line. A creative that raises $\Phi_{\mathcal{R}}$ by attracting a mismatched audience lowers the first-session facets in $\Phi_{\mathcal{A}}$ (Section 6.3), so a higher raw CTR can coexist with a lower rCTR. And because the stages multiply, no amount of reach compensates for an arrival stage near zero (Section 20).

29.4 Three basic propositions

Proposition 29.1 (Fracture and bounds)**.** *For $s \in (0,1]^{28}$ and weights as in* (29.3)*,*

$$\min_j s_j \;\le\; \Phi(s) \;\le\; \sum_{j=1}^{28} w_j s_j, \tag{29.7}$$

and $\Phi(s) \to 0$ whenever any single score $s_j \to 0$ with the others fixed. The same statements hold for each Φ_G.

Proof. Since $\log \Phi = \sum_j w_j \log s_j$ and $\sum_j w_j = 1$, we have $\log \Phi \ge \sum_j w_j \log \min_k s_k = \log \min_k s_k$. Concavity of the logarithm and Jensen's inequality give $\sum_j w_j \log s_j \le \log \sum_j w_j s_j$,

which is the upper bound. If $s_j \to 0$ with $w_j > 0$, then $w_j \log s_j \to -\infty$ while the remaining terms are fixed and finite, so $\log \Phi \to -\infty$. □

A crystal is only as clear as its worst facet allows: an excellent average cannot hide a fractured facet. An arithmetic average of scores would hide exactly that failure.

Proposition 29.2 (Weakest-facet rule)**.** *For each facet, $\partial\Phi/\partial s_j = \Phi\, w_j/s_j$. Moreover, under equal weights, if $s_i < s_k$ and an increment $\Delta > 0$ can be applied to exactly one of the two facets (with $s_k + \Delta \le 1$), applying it to facet i yields the larger clarity.*

Proof. Differentiating $\log \Phi = \sum_j w_j \log s_j$ gives $\partial \log \Phi/\partial s_j = w_j/s_j$, hence the first claim. For the second, with $w_i = w_k = w$, the change in $\log \Phi$ from improving facet $x \in \{s_i, s_k\}$ is $w\, g(x)$ with $g(x) = \log(x + \Delta) - \log x$. Since $g'(x) = 1/(x + \Delta) - 1/x < 0$ for $x > 0$, g is strictly decreasing, so $g(s_i) > g(s_k)$. □

The rule is exact for equal increments, not only to first order. It does not say that the weakest facet is the cheapest to repair; if repair costs differ, compare $w_j/(s_j\, \kappa_j)$ for marginal cost κ_j per unit of score, in the spirit of (17.2).

Proposition 29.3 (Precision of clarity)**.** *Let $\hat{s}_j$ be estimates of s_j with variances σ_j^2, independent across facets. To first order (delta method),*

$$\operatorname{Var}\big(\log \Phi(\hat{s})\big) \approx \sum_{j=1}^{28} \frac{w_j^2\, \sigma_j^2}{s_j^2}. \tag{29.8}$$

Proof. A first-order Taylor expansion of $\log \Phi(\hat{s}) = \sum_j w_j \log \hat{s}_j$ about s gives $\log \Phi(\hat{s}) \approx \log \Phi(s) + \sum_j (w_j/s_j)(\hat{s}_j - s_j)$. Taking the variance of the linear term under independence yields (29.8). □

Uncertainty about clarity is dominated by facets that are both weak and noisily measured, because σ_j^2 is divided by s_j^2. This is the formal role of the evidence group: measurement contracts, mature cohorts, reconciled events and randomized assignment reduce σ_j^2 precisely where (29.8) is most sensitive. Correlated estimates require the full covariance in place of the diagonal sum.

29.5 A synthetic illustration

Figure 14 shows a synthetic facet profile under equal weights. Its overall clarity is $\Phi = 0.504$. The three weakest facets are first-session survival ($s_5 = 0.30$), technical reliability ($s_8 = 0.25$), and the event system ($s_{22} = 0.30$). Adding 0.15 to each raises clarity to $\Phi = 0.528$. Adding the same 0.15 to each of the three strongest facets ($s_1 = 0.80$, $s_{12} = 0.85$, $s_{19} = 0.70$) raises it only to $\Phi = 0.514$, as Proposition 29.2 predicts.

The repair also changes retained click-through. Facets 5 and 8 lie in the arrival group, whose clarity rises from $\Phi_{\mathcal{A}} = 0.374$ to 0.465; facet 22 is an evidence facet and does not enter (29.6). With $\Phi_{\mathcal{R}}$ and $\Phi_{\mathcal{V}}$ unchanged, equation (29.6) implies that rCTR rises by the factor $0.465/0.374 \approx 1.24$, while raw CTR, governed by $\Phi_{\mathcal{R}}$, does not move at all. All scores in this example are invented for illustration; they are not measurements of any game.

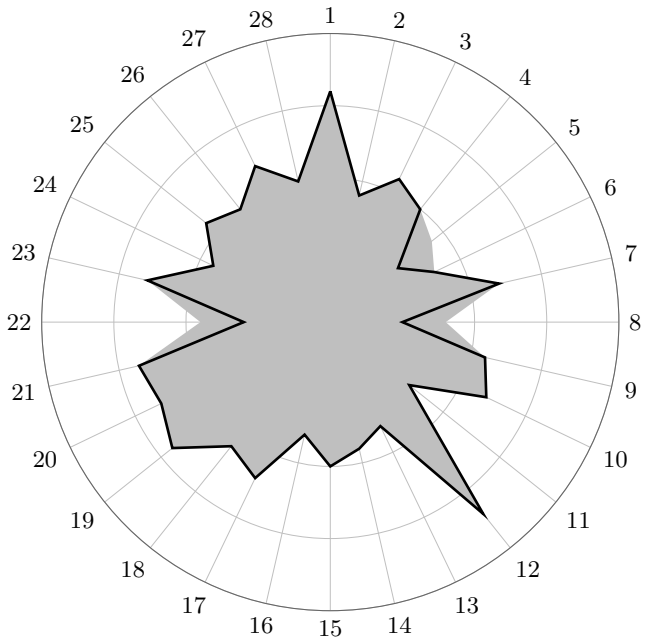


Figure 14. A synthetic crystal profile on 28 radial axes, one per facet, numbered as in Table 7; radial distance is the facet score s_j, with rings at 0.25, 0.50, 0.75 and 1.00. The black line is the synthetic launch profile. The shaded region is the profile after adding 0.15 to each of the three weakest facets (5, 8 and 22); it coincides with the black line on every other axis, so it is visible only as the three bulges at those facets. Scores are illustrative, not measured. Drawn with PGFPLOTS [32] (polar library); scores as listed in Section 29.5.

29.6 Optimal repair under a budget

Proposition 29.2 compares two facets at a time. A studio usually has a fixed budget of effort to spread across many. Let repairing facet j by an amount x_j cost $c_j x_j$, with $c_j > 0$, and let $B > 0$ be the effort available. Allow the objective weights $w_j \geq 0$ to differ from those in (29.3), so that the same result covers overall clarity and retained click-through. Write $J_+ = \{j : w_j > 0\}$.

Relation to classical methods. This is an application of the classical water-filling principle for logarithmic allocation [43]; the proof below specializes it to the stated score bounds and weights. It is not a claim to a new allocation algorithm.

Theorem 29.4 (Water-filling repair)**.** *Consider*

$$\max_{x} \sum_{j=1}^{28} w_j \log(s_j + x_j) \quad \textit{subject to} \quad \sum_{j=1}^{28} c_j x_j \leq B, \qquad 0 \leq x_j \leq 1 - s_j, \tag{29.9}$$

with $s_j \in (0, 1]$ and $B < \sum_{j \in J_+} c_j (1 - s_j)$. The solution is unique. It sets $x_j^* = 0$ for $j \notin J_+$ and

$$x_j^* = \min\Big\{ \Big(\frac{w_j}{\nu c_j} - s_j \Big)^+,\ 1 - s_j \Big\}, \qquad j \in J_+, \tag{29.10}$$

where $\nu > 0$ is any number for which $\sum_j c_j x_j^* = B$. Such ν exist, and x^* does not depend on which is chosen; with equal weights and equal costs ν is itself unique.

Proof. The objective is concave and every constraint is linear, so the Karush–Kuhn–Tucker conditions are both necessary and sufficient [8]. Spending on a facet with $w_j = 0$ uses budget without raising the objective, and the budget binds at the optimum because the objective is strictly increasing in every x_j with $j \in J_+$; hence $x_j^* = 0$ off J_+. For $j \in J_+$, with multiplier $\nu \geq 0$ on the budget and $\alpha_j, \beta_j \geq 0$ on the bounds, stationarity reads $w_j/(s_j + x_j) - \nu c_j + \alpha_j - \beta_j = 0$. If $0 < x_j < 1 - s_j$, then $\alpha_j = \beta_j = 0$ and $s_j + x_j = w_j/(\nu c_j)$; if $x_j = 0$, then $w_j/s_j \leq \nu c_j$; if $x_j = 1 - s_j$, then $w_j \geq \nu c_j$. These three cases are exactly (29.10). The multiplier is positive: $\nu = 0$ would make every $j \in J_+$ satisfy $w_j/(s_j + x_j) + \alpha_j = \beta_j$ with $w_j > 0$, forcing $\beta_j > 0$ and hence $x_j = 1 - s_j$, so spending would equal $\sum_{j \in J_+} c_j(1 - s_j) > B$. The function $g(\nu) = \sum_j c_j x_j^*(\nu)$ is continuous and nonincreasing, equal to zero for $\nu \geq \max_{j \in J_+} w_j/(c_j s_j)$, and tends to $\sum_{j \in J_+} c_j(1 - s_j) > B$ as $\nu \to 0$, so by the intermediate value theorem some $\nu > 0$ satisfies $g(\nu) = B$. The function g can be constant on an interval, so ν need not be unique, but every such ν gives the same allocation: the objective is strictly concave in the coordinates J_+, so its maximizer is unique. With equal weights and costs every coordinate shares one water level, g is strictly decreasing wherever it is positive, and ν is unique. □

Each repaired facet is raised to its own *water level* $w_j/(\nu c_j)$. With equal weights and costs every repaired facet reaches the same level, so the weakest facets are filled first and a facet already above the level is never touched. The marginal criterion that follows Proposition 29.2, $w_j/(s_j \kappa_j)$, is exactly the threshold that decides whether facet j receives any budget: it is repaired if and only if $w_j/(c_j s_j) > \nu$.

Applied to the synthetic profile. Take the profile of Section 29.5, equal costs, and a budget $B = 0.45$, the same total effort as adding 0.15 to each of three facets. Two objectives give two different plans.

1. *Maximize overall clarity* ($w_j = 1/28$). The water level is 0.4071. Seven facets are repaired: 8 by 0.157, facets 5 and 22 by 0.107, facet 11 by 0.057, and facets 6, 13 and 16 by 0.007 each. Clarity rises from 0.504 to 0.528; the three-facet plan of Section 29.5 reaches 0.5276, within 0.1 percent of optimal. Retained click-through rises by a factor of 1.256.

2. *Maximize retained click-through.* By (29.6), $\log \mathrm{rCTR}$ is the clarity objective with $w_j = 1/|G|$ for facets in $\mathcal{R}$, $\mathcal{A}$ and $\mathcal{V}$, and $w_j = 0$ for evidence facets. The whole budget then goes to the arrival group, the smallest and weakest: facet 8 by 0.217, facet 5 by 0.167, and facet 6 by 0.067, all reaching 0.467. Retained click-through rises by a factor of 1.357, against 1.256 for the clarity-optimal plan and 1.245 for the three-facet plan, while overall clarity reaches 0.526.

As a numerical sanity check, each plan was compared with 20,000 random allocations of the full budget, of which about 19,900 were feasible; none did better. The proof, not this check, is the argument. The comparison makes a practical point: aimed at the outcome rather than at the average, the same effort buys a 36 percent gain in retained click-through instead of 26 percent, about 8 percent more retained click-through in absolute terms. A studio should decide which objective it is optimizing before it scores a single facet.

29.7 What the geometry contributes

A fair objection to the crystal is that its content is a weighted geometric mean, and a mean needs no polyhedron. This subsection gives the geometry a precise statistical role and states

how to test whether it earns that role.

Relation to graph learning. Graph-based regularization has established precedents [44]. Here the penalty acts on coefficients attached to facets, rather than on predictions attached to examples. The proposed facet adjacency is a prior whose usefulness requires held-out comparison, not evidence of a new general graph-learning method.

The facet graph. Call two facets adjacent when they share an edge. In the 14-gonal bipyramid each facet shares one edge with each of its two neighbours in the same ring and one with the facet across the girdle, so the facet graph is the prism graph $C_{14}\square K_2$: 28 vertices, each of degree three, and 42 edges, one for each edge of the polyhedron in (29.1). Its Laplacian $L = D - A$, with A the adjacency matrix and $D = 3I$, has eigenvalues

$$\lambda_k = 2 - 2\cos\frac{2\pi k}{14}, \qquad \lambda'_k = 4 - 2\cos\frac{2\pi k}{14}, \qquad k = 0, \dots, 13, \tag{29.11}$$

because the spectrum of a Cartesian product of graphs is the set of pairwise sums of the factor spectra. The eigenvalue zero is simple, the next smallest is $2 - 2\cos(\pi/7) \approx 0.198$, and the largest is 6. A direct numerical eigendecomposition agrees with (29.11) to machine precision.

Estimating 28 weights from a few games. Equation (30.4) has 28 unknown weights, and a studio may have six games or versions to learn from. Ordinary least squares cannot identify all 28 coefficients from so few observations without additional constraints. Let $y \in \mathbb{R}^n$ hold centred log retained click-through and $Z \in \mathbb{R}^{n\times 28}$ the centred log facet scores, and penalize disagreement between adjacent facets:

$$\hat{w} = \arg\min_{w} \big\{ \|y - Zw\|^2 + \tau\, w^\top L w \big\}, \qquad w^\top L w = \sum_{j\sim k} (w_j - w_k)^2, \qquad \tau > 0. \tag{29.12}$$

Proposition 29.5 (Uniqueness of the regularized estimate)**.** *If $\tau > 0$ and $Z\mathbf{1} \neq 0$, then $Z^\top Z + \tau L$ is positive definite, and* (29.12) *has the unique solution $\hat{w} = (Z^\top Z + \tau L)^{-1} Z^\top y$ for every $n \geq 2$.*

Proof. Both terms are positive semidefinite, so it suffices to show that $v^\top (Z^\top Z + \tau L) v = \|Zv\|^2 + \tau v^\top L v = 0$ forces $v = 0$. Both terms must vanish. Since the facet graph is connected, $v^\top L v = 0$ only when $v = c\mathbf{1}$ for some scalar c. Then $Zv = cZ\mathbf{1} = 0$, and $Z\mathbf{1} \neq 0$ gives $c = 0$. □

This is uniqueness of a penalized estimator, not identification of the true 28 weights from the observations. For $n < 28$, the data alone generally leave unobserved directions; the penalty supplies the choice along those directions. Equation (29.12) also imposes neither nonnegativity nor a sum-to-one constraint, so its fitted coefficients are not automatically the normalized clarity weights of (29.3). Predictive validity and sensitivity to τ require out-of-sample evaluation.

After centring, $Z\mathbf{1} = 0$ exactly when every game has the same mean log facet score, so the condition only asks that not all games share that mean; with a single game, centring makes $Z = 0$ and the condition fails, which is why $n \geq 2$ is required. In the eigenbasis of L, the component of w along the eigenvector for λ_k is penalized by $\tau\lambda_k$: weight patterns that vary slowly around the crystal are almost free, and patterns that alternate between neighbours are penalized up to 6τ. The geometry is therefore a prior: neighbouring structures are expected to matter similarly.

When the geometry should be dropped. That prior depends on which facets are neighbours. The default labeling follows the manuscript, in which adjacent sections treat related mechanisms, and any other labeling must be declared before outcomes are examined. The geometry earns its place only if it predicts better than a geometry-free alternative. The test is direct: replace L by the identity, which gives ordinary ridge regression, and compare held-out prediction error across games. If ridge does as well, the polyhedron adds nothing beyond its value as a diagram, and it should be presented as one.

29.8 Limits and a falsifiable prediction

The crystal inherits every caution of the manuscript. Scores are only as good as their contracts; weights are value judgements; the group assignment is declared, not discovered; and the facet labeling, which fixes the adjacency prior of Section 29.7, is declared rather than estimated. The multiplicative form (29.5) is the substantive assumption. It encodes complementarity, as in (17.4), and it can be wrong: some structures may substitute for one another rather than multiply.

The model therefore makes a testable claim. Under randomized assignment (Section 14), an improvement that raises the weakest arrival facet should increase mature rCTR more than an equal-effort improvement to an already strong facet in the same group. This prediction additionally requires comparable score gains, costs, weights, and no unmodeled spillovers. Equal engineering effort need not produce equal score increments. Repeated, precise contrary results would challenge that operational prediction, but would not by themselves identify an additive alternative or disprove every multiplicative model. Section 31 sets out a proposed prospective version of this test on Push Party, with its sample size.

Practical implication. Score all 28 facets before buying more impressions. Judge acquisition by retained click-through, not raw click-through, and repair the weakest facet that the evidence group can measure reliably. A brilliant thumbnail on a fractured crystal still loses its light.

30 Beyond one game: extensions and open problems

The preceding sections were built around Roblox and around one studio's evidence. Most of the machinery does not depend on either. This section states where the framework extends, what changes when it does, and which questions remain open. It is deliberately exploratory: each extension is a direction with a stated first model, not a finished result.

30.1 Other creator platforms

Any platform where independent creators publish experiences into a shared discovery system, such as Fortnite Creative, Minecraft servers, or Rec Room, has the same skeleton: impressions become clicks, clicks become first sessions, first sessions become returning players, and a platform-controlled distribution layer sits between the creator and the audience. The cohort renewal equation (3.1), first-session survival (Section 5), the multistate return model (Section 4), measurement contracts (Section 2), and the experimental designs of Section 14 transfer without change, because they describe players and evidence rather than any one platform.

What does not transfer is everything platform-specific: how discovery ranks experiences, which analytics a creator can see, how payments and revenue sharing work, which advertising tools exist, and which events are logged by default. In the notation of Section 29, the facet

structures transfer while the facet *proxies* of Table 7 must be rebuilt for each platform. Any platform-specific statement in this manuscript should be re-verified against that platform's current documentation before use.

30.2 Live-service games and subscription apps

Outside creator platforms, a mobile free-to-play game, a live-service PC title, or a subscription app faces the same accounting. Daily activity is acquisition convolved with a return kernel; revenue is retained value net of acquisition cost; updates act on transition rates rather than only on attention. The main differences are scale and control. A large publisher can randomize at the level of individual users across millions of installs, which tightens the inference of Section 15 considerably, but it also faces paid acquisition markets whose saturation can be far steeper than the small-campaign response behind equation (12.7). The framework applies; the parameter regime is different.

30.3 From one game to a studio portfolio

A studio that operates several games must decide where the next hour of work goes. Let game $g \in \{1, \dots, G\}$ return expected value $V_g(h_g)$ from h_g hours, with each V_g increasing and concave, and let H be the hours available. The portfolio problem has the same form as the budget allocation (12.5), with hours across games in place of dollars across creatives:

$$\max_{h_1,\dots,h_G \ge 0} \sum_{g=1}^{G} V_g(h_g) \quad \text{subject to} \quad \sum_{g=1}^{G} h_g \le H. \tag{30.1}$$

The Karush–Kuhn–Tucker conditions [8] give, for a multiplier $\lambda \ge 0$,

$$V_g'(h_g^*) = \lambda \ \text{ if } h_g^* > 0, \qquad V_g'(0) \le \lambda \ \text{ if } h_g^* = 0. \tag{30.2}$$

Every funded game earns the same marginal return on its last hour, and a game whose *first* hour is worth less than that common return receives none. This is the portfolio form of the rational stop in Definition 1.1: retiring a game is not a failure of the model but one of its outputs. The open difficulty is that each V_g is uncertain and estimated from cohorts of very different maturity, so the allocation should be re-solved as the evidence of Section 16 matures rather than fixed at launch.

30.4 Explore and exploit in creative testing

Section 12 treated advertising creatives as fixed alternatives to compare. In practice a studio tests creatives while it spends, and each impression spent on a weaker creative is an impression not spent on a stronger one. This is the explore–exploit trade-off of the multi-armed bandit problem [9].

Thompson sampling [35] handles the trade-off directly. For creative i with n_i impressions and r_i *retained* conversions (not clicks), keep the posterior of Section 15,

$$\theta_i \mid \mathcal{D} \sim \mathrm{Beta}(1 + r_i,\ 1 + n_i - r_i), \tag{30.3}$$

draw one sample $\tilde{\theta}_i$ from each posterior, and assign the next slice of budget to the creative with the largest draw. Creatives with little evidence keep wide posteriors and are still explored;

creatives with strong evidence are exploited. For Bernoulli rewards observed immediately, the expected regret of Thompson sampling grows only logarithmically in the number of impressions [36].

The Roblox setting breaks that last assumption in an instructive way. The reward that matters, retained conversion, is observed only after the return window closes (Section 16), so the bandit learns with a delay of days. Two practical responses follow. The first uses a fast proxy, such as five-minute first-session survival, and switches to the matured outcome as cohorts close. The second keeps a fixed exploration floor so that no creative is starved before its evidence matures. Optimizing raw click-through instead of retained conversion turns the bandit into an efficient machine for finding the creative that attracts the wrong players, which is exactly the failure Section 29.3 warns against.

Prior work and scope. Delayed-feedback learning is an established problem [45]. The comparison below illustrates a target mismatch in a specific synthetic environment; it neither introduces a new delay-aware algorithm nor establishes superiority over the delayed-feedback literature.

A simulation study. To measure how much these choices matter, consider four creatives whose click-through rates are the observed campaign values of Table 5: 4.7, 2.3, 2.0 and 1.7 percent. Their D7 retention per click is invented, and chosen so that the highest-click creative attracts the worst-matched players: 1.0, 3.5, 2.5 and 2.0 percent. Per impression, creative B is then the best by a wide margin, and creative A, the click winner, is only third. A campaign buys 10,000 impressions a day for 28 days and sets each day's split at the start of the day, as an advertiser setting daily budgets would. Retention matures seven days after the impression. Every Thompson policy keeps an exploration floor of 2.5 percent per creative. Five policies are compared:

1. a uniform split;
2. Thompson sampling on clicks;
3. Thompson sampling on matured retention, splitting uniformly until the first outcomes mature;
4. Thompson sampling on an *aligned* proxy, five-minute survival, which per impression ranks all four creatives in the same order as retention (B, C, A, D), switching to retention when it matures;
5. the same with a *misaligned* proxy, which per impression ranks creative A first because it inherits A's click volume.

Loss is measured in expectation as retained players forgone relative to spending every impression on the best creative, which removes outcome noise and leaves only the noise of the policy's choices. A Beta$(1, 1)$ prior and 400 posterior draws per daily allocation define the Thompson policies. The oracle has no exploration floor and is an unattainable benchmark, so even a well-learned exploratory policy retains positive regret. Each non-oracle policy was run 2,000 times; Table 8 reports means.

Three results stand out. First, optimizing clicks is worse than not optimizing at all: Thompson sampling on clicks converges confidently on creative A and loses 18 percent more retained

Table 8. Simulated creative testing over 28 days at 10,000 impressions per day, 2,000 replications per policy. Standard errors of the mean loss are 0.002 for the first two rows and at most 0.27 for the rest. All retention values are synthetic; click-through rates are the observed campaign values of Table 5.

Policy	Retained	Lost vs. best	Share of oracle	Budget on best
Uniform split	148.1	77.3	65.7%	25.0%
Thompson on clicks	133.8	91.6	59.4%	3.3%
Thompson on retention, delayed	181.4	44.0	80.5%	56.0%
Aligned proxy, then retention	186.7	38.7	82.8%	61.1%
Misaligned proxy, then retention	177.9	47.5	78.9%	51.9%
Oracle (always best)	225.4	0.0	100.0%	100.0%

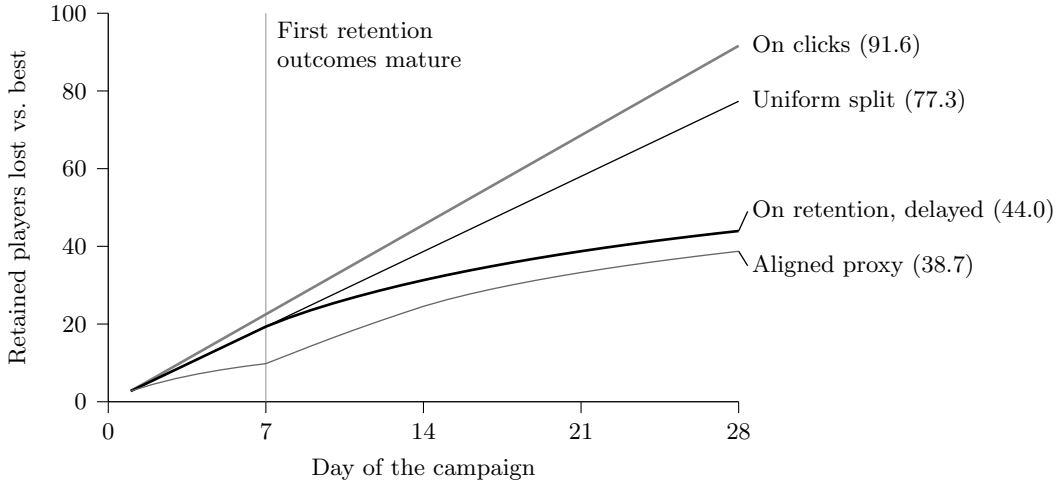


Figure 15. Cumulative retained players lost relative to always choosing the best creative, averaged over 2,000 runs of each policy: Thompson sampling on clicks (heavy gray line, "on clicks"), uniform split (thin black line), Thompson sampling on retention with a seven-day delay (heavy black line, "on retention, delayed"), and an aligned proxy followed by retention (thin gray line, "aligned proxy"). Final losses are given in parentheses. Before day 7 the delayed policy has no retention data and splits uniformly, so its line follows the uniform line; the aligned proxy separates from it at once. Drawn with PGFPLOTS [32]; simulated with NumPy [33].

players than a uniform split. Second, learning from the right outcome pays even when that outcome arrives a week late: delayed Thompson sampling on retention cuts the loss by 43 percent relative to the uniform split. Third, a proxy helps only if it ranks creatives the way the target does. The aligned proxy cuts the remaining loss by a further 12 percent; the misaligned proxy, which looks equally reasonable on paper, raises it by 8 percent. The practical rule is to validate a proxy's *ranking* of past creatives against matured retention before letting it steer spending.

The numbers depend on the invented retention rates, and the size of each effect will differ in a real campaign. These policy rankings are specific to the declared rates, horizon, prior, and exploration floor; they are not a uniform guarantee across environments. A click-based objective can select a different creative from a retained-conversion objective.

30.5 Extending the crystal

Three extensions of the crystal model are natural.

Estimated weights. Taking logarithms of equation (29.6) gives a model that is linear in its unknowns. Across games or versions g,

$$\log \mathrm{rCTR}_g = \beta_0 + \sum_j w_j \log s_{gj} + \varepsilon_g, \tag{30.4}$$

so coefficients can be fitted by least squares when the design has adequate rank. The model is an assumption to test, not an identity that validates its fitted coefficients. With twenty-eight facets and a fitted intercept, unique unpenalized estimation requires at least twenty-nine observations and a full-column-rank design, which a single studio rarely has. Two remedies exist: estimate one weight per facet group, three in the rCTR equation, or use the adjacency-regularized estimator of Proposition 29.5, which is unique for any two or more games that differ in their mean log facet score. Its uniqueness comes partly from the penalty and does not identify the true weights from a small sample. Hyperparameters and facet labeling should be fixed using training data only; held-out games must remain untouched until evaluation. Pooling evidence across studios helps either way.

A dynamic crystal. Facet scores are not fixed. Novelty fades, technical debt accumulates, and economies inflate. A first model is

$$s_j(t+1) = s_j(t) + \eta_j u_j(t) - \delta_j s_j(t), \tag{30.5}$$

where $u_j(t)$ is effort directed at facet j, η_j its effectiveness, and δ_j the facet's decay rate. Under constant effort the score settles at $s_j^* = \eta_j u_j / \delta_j$ when this value lies in $(0, 1]$; otherwise bounded scores require a saturating or projected dynamic. A facet with fast decay needs continuous investment merely to hold its value. This connects the crystal to the update cadence of Section 26: an update schedule is, in this model, a maintenance schedule for the fastest-decaying facets.

Genre crystals. The partition of Table 7 was declared for a round-based social game. A tycoon, a horror experience, or a roleplay world may assign the same structures to different groups or weight them very differently. Comparing estimated crystals across genres is itself a research question.

30.6 Open problems

The following questions are stated so that each could be answered by a well-designed study.

1. Does repairing the weakest arrival facet raise mature retained click-through more than an equal-effort repair of a strong facet, as Proposition 29.2 predicts? (Section 29.8 gives the test.)
2. How much of the variation in retained click-through across games do the three group clarities explain, and are the estimated weights in equation (30.4) stable across genres?
3. What is the shortest proxy outcome that predicts matured D7 retention well enough to drive the delayed bandit of Section 30.4 without steering it toward the wrong audience?
4. How should experiments be designed when treatment effects spread through friend networks across servers, extending the interference designs of Section 14?

5. At what decay rates in equation (30.5) does a fixed release cadence stop being sufficient, so that the cadence of Section 26 must adapt to measured decay?
6. Which facet structures transfer across creator platforms unchanged, and which must be rebuilt, when the same game is published on two platforms?

Practical implication. The framework is not limited to one platform or one game. Treat a portfolio as one allocation problem, test creatives against retained outcomes rather than clicks, and expect every facet to decay without maintenance. The open problems above are where the next evidence should come from, and some of them can be answered by a single studio with a well-instrumented release.

31 A proposed prospective test on Push Party

The empirical claims in Section 29 remain unvalidated. This section proposes a prospective test for Push Party; it is not a completed preregistration. Eligibility, interventions, estimands, randomization, stopping rules, and analysis must be frozen and timestamped before enrollment. Any required ethics, privacy, consent, and platform review must also be resolved first. It is written to be run by a small studio with ordinary platform tools, and to be worth reporting whatever it finds.

31.1 Hypotheses

Let C be the current game, W a version that repairs the weakest arrival facet, and S a version that improves the strongest arrival facet with the same engineering effort.

H1. (Primary.) Exact-day D1 retention is higher under W than under S. This is the falsifiable prediction of Section 29.8: equal effort buys more when it is aimed at the weakest facet.

H2. (Secondary.) D1 under W exceeds D1 under C.

H3. (Secondary.) Five-minute first-session survival $S(5)$ under W exceeds that under C.

A nonsignificant H1 result is inconclusive unless its interval excludes a prespecified effect of practical interest. A precise reversal challenges this particular intervention prediction. Equal engineering effort is not equal score improvement, so no single arm comparison can establish or reject the entire multiplicative model.

31.2 Arms and assignment

Before building either change, the four arrival facets (5, 6, 8 and 23) are scored with the proxies of Table 7 on the most recent mature cohorts. The lowest-scoring facet defines W and the highest-scoring defines S. Engineering hours for W and S are logged and capped at the same total, so that the comparison is between targets, not between amounts of work. If the scores point where the launch audit suggests, W is an action-first first round in which a new player completes a meaningful push within thirty seconds, and the five-minute survival of 12.43 percent in Table 3 is its baseline.

Each new user is assigned at first join, with equal probability, by a prespecified salted pseudo-random hash of the user identifier and study key, so that assignment is fixed on every later visit and independent of acquisition source. Assignment, version and arm are logged with the events

of Section 22. Because the change affects a player's first rounds rather than shared server state, assignment is per player; spillover through mixed servers is assessed using server arm composition. This diagnostic cannot remove interference bias. If either intervention changes shared play materially, use server-level assignment and recompute power for clustering before enrollment; the individual-user calculations below then do not apply. The test changes only game design, within the studio's normal update practice; it collects nothing beyond platform analytics and the event schema of Section 22, and it follows Roblox's terms for experiences played by minors.

31.3 Sample size

For two proportions p_0 and p_1, a two-sided test at level α with power $1-\beta$ needs, per arm,

$$n = \frac{\left(z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta}\sqrt{p_0(1-p_0)+p_1(1-p_1)}\right)^2}{(p_1-p_0)^2}, \qquad \bar{p} = \frac{p_0+p_1}{2}. \tag{31.1}$$

With power 0.80, the October 22 cohort's D1 of 1.53 percent as the baseline, and the five-minute survival of 12.43 percent:

Outcome and detectable change	Per arm, $\alpha = 0.05$	Per arm, $\alpha = 0.025$
D1: 1.53% to 2.50%	3,293	3,988
D1: 1.53% to 3.06% (doubling)	1,503	1,820
$S(5)$: 12.43% to 15.00%	2,812	3,405
$S(5)$: 12.43% to 16.00%	1,501	1,818

Use hierarchical testing: test H1 two-sided at 0.05; only after rejecting H1 in the predicted direction, test H2 and H3 with Holm at familywise 0.05. Otherwise report secondary contrasts descriptively. The secondary family's first Holm step uses $\alpha/2$, so its planned contrasts are sized from the right-hand column. The proposed individual-user design targets 4,000 new users per arm, 12,000 in total. These are normal-approximation calculations assuming independent outcomes and complete follow-up. For H1, S is provisionally assumed to retain 1.53 percent; this is not an observed baseline for the unbuilt S intervention. Re-estimate the required sample before registration if that assumption changes. At that size the primary test of H1, at $\alpha = 0.05$, detects a D1 change from 1.53 to 2.40 percent, and the Holm-adjusted secondary tests detect D1 from 1.53 to 2.50 percent and five-minute survival from 12.43 to 14.8 percent, each with power 0.80. At the launch acquisition rate of about 5,400 new users a day (Table 4), enrollment takes between two and three days. At the far lower organic rate after paid acquisition stopped it will take proportionally longer, and a short paid burst under the stopping rules of Section 20 is an acceptable way to reach the sample, provided its traffic is split across arms by the same hash.

31.4 Analysis plan

The sample size is fixed in advance and there is no interim test of the primary outcome; outcome-driven interim changes are not permitted under this fixed-sample plan. Any sequential alternative must be specified and powered before enrollment. Safety monitoring is separate; a safety stop is reported and does not inherit the nominal power or fixed-sample inferential guarantee. Each contrast is reported as a difference in proportions with a Newcombe interval [37], together with the raw counts, the number of users excluded, and the reason for each exclusion. Assignment defines the intention-to-treat population. Fix bot and duplicate exclusions without

examining outcomes; departures and non-returners remain in their assigned arm. A complete D1 window without a return event counts as zero only if event capture is verified; telemetry loss is missing and requires reported bounds or sensitivity analysis. H2 and H3 follow the hierarchical gate and a Holm correction [38]. D1 is analyzed only after every enrolled user's day-one window has closed (Section 16); D7 is reported as exploratory.

Three guardrails stop the test early for harm: a rise in crash or error rate above its pre-test level by more than half, any increase in failed purchase receipts, and any change to prices or monetization during enrollment, which would confound the arms. The frozen protocol, with a cryptographic hash of this section's source, must be posted publicly with a timestamp before the first user is enrolled; no such posting is asserted here.

31.5 What each result would mean

If W beats S, the weakest-facet prediction survives its first test and the crystal gains evidence it currently lacks. If W and S are indistinguishable, report the interval: failure to reject does not establish equivalence or exclude a one-percentage-point effect unless the interval does so. If S beats W precisely, reconsider intervention fidelity, relative score gains, costs, and interference as well as the model. That reversal alone does not establish substitution or justify an additive replacement. Each outcome is informative, which is the reason to run the test.

Practical implication. Score the arrival facets, build the two matched changes, and randomize by user. The proposed twelve-thousand-player design could test this operational question under its assumptions; clustering, missingness, or a different baseline can require more players. Publish the result whichever way it falls.

32 Conclusion

This manuscript has treated a Roblox experience as a partially observed, controlled stochastic system. Its contribution is a set of connected instruments rather than a single formula: measurement contracts that make numbers comparable; cohort renewal and multistate models that separate acquisition from return; survival calculus that locates early friction; causal designs that respect changing audiences and social interference; conservation laws for economies; constrained allocation of advertising and production hours; and a theory of updates that values durable transitions over launch-day pulses. The crystal model gathers these 28 structures into one object and links them to retained click-through, and Section 30 carries the framework beyond a single game, to other platforms, studio portfolios, and adaptive creative testing.

Three findings recur throughout. First, visible metrics are coupled outcomes, so a change in one number rarely identifies its cause without a declared estimand and a comparison. Second, early and structural improvements, such as reliable first loops, readable choices, and preserved progression, propagate value to every later state, as the hazard functional derivative (5.8) and the adjoint sensitivity (18.12) make explicit. Third, durable success comes from updates that extend a recognizable core grammar with new situations, rather than from reach alone.

Three applications make the framework concrete; they are not claims to invention of their underlying methods. Theorem 29.4 applies classical logarithmic allocation [43] to the weakest-facet model and shows that the objective, retained click-through or average clarity, changes where effort should go. Proposition 29.5 gives the crystal's geometry a testable regularization role; it does not establish data-only identification of the true weights. The simulation study of

Section 30.4 quantifies a mistake that is easy to make and expensive: steering ad spend by clicks when the outcome that matters arrives a week later.

The work has clear limits. The launch audit is a single provisional case, the synthetic experiments verify model implications rather than player behaviour, and the crystal model has not yet been tested against data. Section 31 proposes a test with about twelve thousand new players and three arms, subject to its stated power and interference assumptions. A timestamped protocol and any required review must precede enrollment. Running it would provide evidence about a particular operational prediction, whichever way the result falls.

References

[1] M. A. Hernan and J. M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020; online edition maintained by the authors. https://miguelhernan.org/whatifbook.

[2] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958. https://doi.org/10.1080/01621459.1958.10501452.

[3] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34(2):187–202, 1972. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

[4] J. D. C. Little. A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3):383–387, 1961. https://doi.org/10.1287/opre.9.3.383.

[5] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. https://doi.org/10.1093/biomet/58.1.83.

[6] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. https://arxiv.org/abs/1810.08240.

[7] M. Dudik, J. Langford, and L. Li. Doubly robust policy evaluation and learning. *Proceedings of ICML*, 2011. https://arxiv.org/abs/1103.4601.

[8] J. Nocedal and S. J. Wright. *Numerical Optimization*, second edition. Springer, 2006. https://doi.org/10.1007/978-0-387-40065-5.

[9] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*, second edition. MIT Press, 2018. https://mitpress.mit.edu/9780262039246/reinforcement-learning/.

[10] Roblox. Analytics. Creator Hub documentation. https://create.roblox.com/docs/production/analytics.

[11] Roblox. Retention. Creator Hub documentation. https://create.roblox.com/docs/production/analytics/retention.

[12] Roblox. Engagement. Creator Hub documentation. https://create.roblox.com/docs/production/analytics/engagement.

[13] Roblox. Discovery. Creator Hub documentation. https://create.roblox.com/docs/discovery.

[14] Roblox. Discovery FAQ. Creator Hub documentation. https://create.roblox.com/docs/discovery-faq.

[15] Roblox. Ads Manager. Creator Hub documentation. https://create.roblox.com/docs/production/promotion/ads-manager.

[16] Roblox. Funnel events. Creator Hub documentation. https://create.roblox.com/docs/production/analytics/funnel-events.

[17] Roblox. Economy events. Creator Hub documentation. https://create.roblox.com/docs/production/analytics/economy-events.

[18] Roblox. Event types. Creator Hub documentation. https://create.roblox.com/docs/production/analytics/event-types.

[19] Roblox. Performance optimization. Creator Hub documentation. https://create.roblox.com/docs/performance-optimization.

[20] Roblox. Test on hardware. Creator Hub documentation. https://create.roblox.com/docs/performance-optimization/test-on-hardware.

[21] Roblox. Data store error codes and limits. Creator Hub documentation. https://create.roblox.com/docs/cloud-services/data-stores/error-codes-and-limits.

[22] Roblox. Matchmaking. Creator Hub documentation. https://create.roblox.com/docs/matchmaking.

[23] Roblox. Developer products. Creator Hub documentation. https://create.roblox.com/docs/production/monetization/developer-products.

[24] Roblox. Paid random items policy guidelines. Creator Hub documentation. https://create.roblox.com/docs/production/monetization/paid-random-items.

[25] Roblox. Experience events and updates. Creator Hub documentation. https://create.roblox.com/docs/production/promotion/experience-events.

[26] Roblox. Release game updates. Creator Hub documentation. https://create.roblox.com/docs/projects/update-games.

[27] Roblox. Experience configs. Creator Hub documentation. https://create.roblox.com/docs/production/configs.

[28] Roblox. Creator Rewards. Creator Hub documentation. https://create.roblox.com/docs/creator-rewards.

[29] Nikilis. Murder Mystery 2. Official Roblox experience page, place 142823291. https://www.roblox.com/games/142823291/Murder-Mystery-2.

[30] N. Dasari. Push Party: Full Analytics Audit. Author-authorized Creator Dashboard and Ads Manager extraction, October 26, 2026. Unpublished descriptive data package; provisional attribution. Universe 8164718192. Captured 23:35–23:55 UTC, with later campaign-state checks.

[31] T. Tantau. *The TikZ and PGF Packages: Manual for Version 3.1.10*. 2023. https://ctan.org/pkg/pgf.

[32] C. Feuersänger. *Manual for Package PGFPLOTS, Version 1.18.1*. 2021. https://ctan.org/pkg/pgfplots.

[33] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. Array programming with NumPy. *Nature*, 585:357–362, 2020. https://doi.org/10.1038/s41586-020-2649-2.

[34] P. Virtanen, R. Gommers, T. E. Oliphant, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020. https://doi.org/10.1038/s41592-019-0686-2.

[35] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. https://doi.org/10.1093/biomet/25.3-4.285.

[36] S. Agrawal and N. Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, JMLR Workshop and Conference Proceedings 23:39.1–39.26, 2012. https://proceedings.mlr.press/v23/agrawal12.html.

[37] R. G. Newcombe. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17(8):873–890, 1998. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I.

[38] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. https://www.jstor.org/stable/4615733.

[39] M. Viljanen, A. Airola, J. Heikkonen, and T. Pahikkala. Playtime measurement with survival analysis. *IEEE Transactions on Games*, 10(2):128–138, 2018. https://doi.org/10.1109/TCIAIG.2017.2727642. Author preprint: https://arxiv.org/abs/1701.02359.

[40] K. Rothmeier, N. Pflanzl, J. A. Hüllmann, and M. Preuss. Prediction of player churn and disengagement based on user activity data of a freemium online strategy game. *IEEE Transactions on Games*, 13(1):78–88, 2021. https://doi.org/10.1109/TG.2020.2992282.

[41] M. Viljanen, A. Airola, J. Heikkonen, and T. Pahikkala. A/B-test of retention and monetization using the Cox model. *Proceedings of AIIDE*, 13(1):248–254, 2017. https://doi.org/10.1609/aiide.v13i1.12941.

[42] J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: network exposure to multiple universes. *Proceedings of KDD*, 329–337, 2013. Author preprint: https://arxiv.org/abs/1305.6979.

[43] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. Section 5.5.3, Example 5.2. https://stanford.edu/~boyd/cvxbook/.

[44] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006. https://www.jmlr.org/papers/v7/belkin06a.html.

[45] P. Joulani, A. Gyorgy, and C. Szepesvari. Online learning under delayed feedback. *Proceedings of ICML*, PMLR 28(3):1453–1461, 2013. https://proceedings.mlr.press/v28/joulani13.html.

Figures and software. All fifteen figures use vector graphics. They are drawn with PGFPLOTS 1.18.1 [32] and TikZ/PGF 3.1.10 [31], compiled with pdfTeX (TeX Live 2023). No figure is a generated or raster image. The simulation study of Section 30.4 uses a fixed random seed;

reproduction also depends on the numerical-library versions. Plotted values were computed in Python 3.12.14 with NumPy 2.3.5 [33] and SciPy 1.17.0 [34]: survival and cohort curves by direct evaluation of the stated formulas and discrete convolution, the optima in Figures 7 and 10 by Brent's root-finding method on the first-order conditions, the isolines in Figure 11 in closed form from the characteristic polynomial of B_τ (verified against numerical eigenvalues to 10^{-12}), and the crystal of Figure 13 by orthographic projection of the 14-gonal bipyramid with back-face removal. Figure 8 uses the audit values exactly as rendered in the source dashboard. The complete LaTeX source, the figure-generation script, and the analysis scripts for Sections 29.6, 30.4 and 31, which provide the formulas, figure-generation code, synthetic parameters, and analysis procedures, are published at nishanthdasari.com (source package).

Source and calculation note. Platform documents and the public experience description were inspected on October 26, 2026. Their contents can change. All synthetic figure parameters are given in the relevant section or caption. The numerical examples do not use private Murder Mystery 2 telemetry. The audit's event counts and cohort percentages are preserved as reported rather than treated as causal estimates.