

Whitepaper (in partnership with Amplitude)

THE FRANKENSTEIN STACK PROBLEM:

Why orgs are switching to Amplitude Experiment for a unified approach?

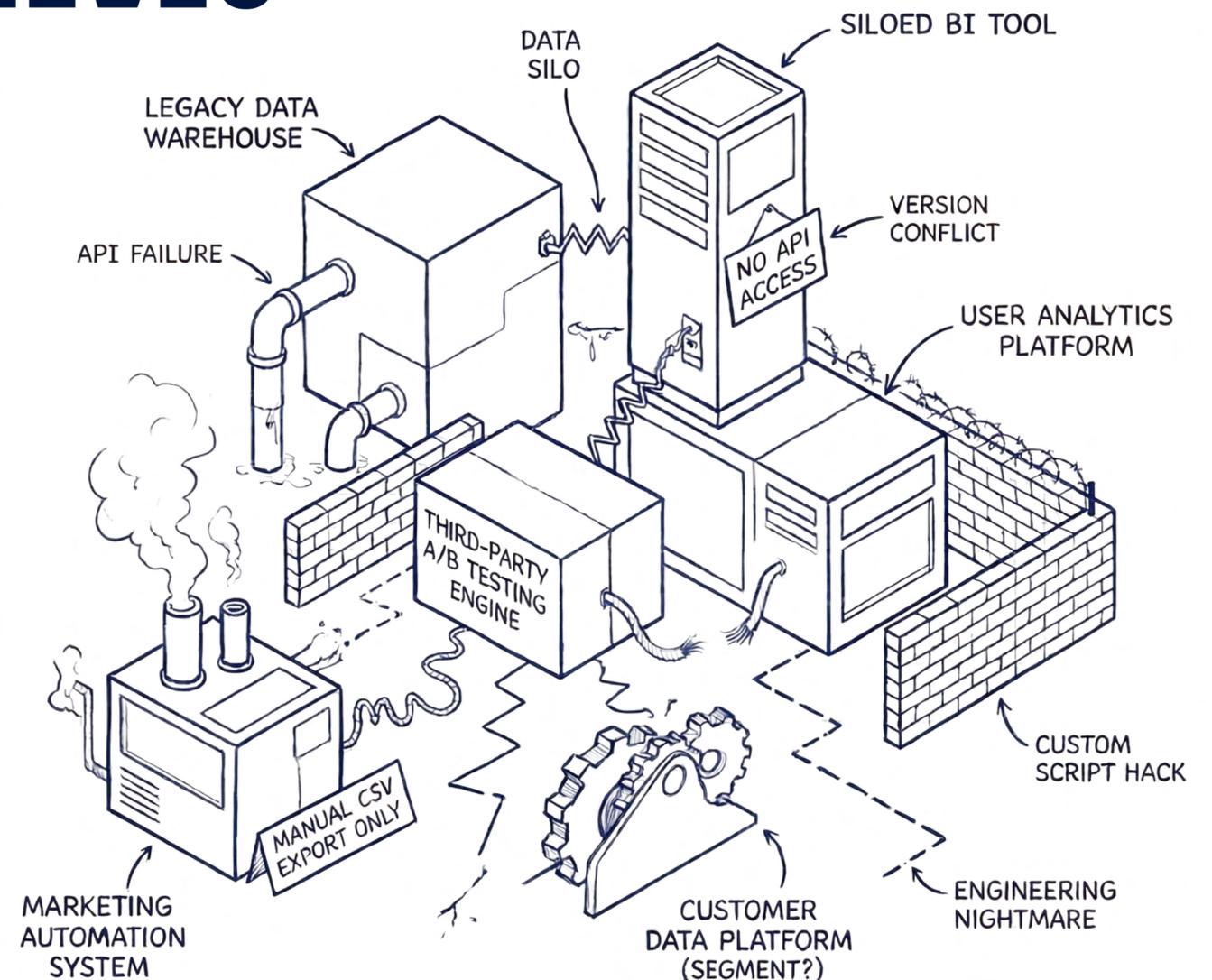


TABLE OF CONTENTS:

2: Foreward

4: About Speero

5: Why Disconnected Stacks Fail

7: A Unified Approach

8: Solve the Identity Crisis

10: Advanced Targeting via JSON Payloads

12: Integrate into the CMS and Product

13: Unlock the Power of Unified Experimentation

14: The Velocity Unlock

15: The Trust Unlock

16: The Insights Unlock

17: Optimize the System, Not the Tool

18: Power Experimentation with Amplitude

FOREWORD

You cannot build a high-maturity experimentation program on what I call the "Frankenstein Stack"—a cobbled-together mess of isolated analytics tools, testing engines, and data platforms that don't talk to each other.

And yet, too many companies still use separate A/B testing and feature flagging tools that are disconnected from customer behavior and analytics. The result?

- Separate data silos that force teams to rely on external analytics to understand outcomes.
- Scattered, tactical experiments that are rarely tied to real business impact.
- Unreliable results that erode trust and confidence in experimentation.
- An inability to take on high-leverage bets that actually move the needle.
- Glacial experimentation timelines, crawling decisions, and missed opportunities.

A better alternative? One connected platform for every step of the experimentation process—from generating a hypothesis and targeting users to measuring results and understanding which experiences users receive.

Unifying your data models, analysis, and testing workflows enables you to design, ship, and see results faster than ever, quantify your impact with precision, and continuously innovate.

(Continued on next page)

The whitepaper was authored by Jonny Longden, though the foundational insights were driven by Eddie Aguilar and Ben Labay, who conducted the primary interviews and compiled the strategic point of view.

FOREWORD (Continued)

That's why more and more teams are saying goodbye to the "Frankenstein Stack" and hello to one connected system for experimentation.

To understand the how and why behind this shift, I anonymously interviewed four organizations that have switched to Amplitude for a unified approach to experimentation:

- [Huckleberry Labs](#)
- [Zwift](#)
- [FanDuel](#)
- [BigCommerce](#)

These teams stopped treating experimentation and analytics as separate disciplines and combined them in a single source of truth.

I also interviewed Eric Metelka, Director of Product Management for Experimentation at Amplitude, for his perspective on the current state of experimentation and Amplitude's vision for a more connected approach.

"I've seen teams lose trust in their experimentation programs and processes. That's why orgs need a foundation of high data rigor when they begin experimenting."—Eric Metelka

Let's explore how merging your data foundation unlocks the velocity and trust you've been chasing.

Jonny Longden

Chief Growth
Officer, Speero



ABOUT SPEERO

We build, manage, and scale growth experimentation for modern product and marketing teams.

Powered by UX research, data & engineering, AB testing, and strategic operating systems (XOS).

Forget isolated A/B tests and half-baked optimization.

We help you build a **true operating experimentation system** that replaces gut calls and guesswork with data, speed, and clarity.

No more siloed teams, slow rollouts, or output faking as progress.

We embed a **high-velocity framework** that turns fragmented initiatives into repeatable ones, creating an engine for growth.

miro



NATIVE



DECATHLON

Gympass

WHY DISCONNECTED STACKS FAIL (A.K.A The Law of Most Surprises)

If you are running a growth program, you've probably seen this scenario:

1. You run a test in a third-party tool
2. The testing tool says you won
3. Your analytics tool says you lost
4. The data science team shows polluted samples
5. You can't decide what to do next
6. Trust in the program erodes

You may think you have a people problem. You don't. You have a system silo and data integrity problem.

The team at **Huckleberry Labs** (a brand with complex user journeys involving parents, children, and multiple devices) encountered this and internally dubbed it "**the law of most surprises.**"

"We had an internal thing that we called the law of most surprises. With [our prior tool], every time we did something, there'd be a surprise. It would say 'Yes, you can do that,' but then we'd determine that the result only applied to 70% of people because of some other factor." —Director, Business Analytics, Huckleberry Labs

If you've been in the UX industry for a while, you probably know the "law of least surprises": *You want to guide users through the flow without jarring interruptions.*

But for Huckleberry Labs, their prior stack and disparate tools made **every test an unwanted surprise.**

They were constantly sampling data, so analysts were always unsure if they were seeing the full data set.

This forced them to manually stitch data in their prior tools, in a *"very manually intensive process."*

WHY DISCONNECTED STACKS FAIL (Continued)

They would launch a test only to find a bug mid-flight, or that they had to patch a unique issue they didn't see coming. When your data lives in silos, you spend more time validating the tool than validating the hypothesis.

For BigCommerce, this friction was even more visible. They ran client-side testing via Optimizely while trying to reconcile that data with Tableau and Amplitude. This ended with **flicker issues** and, even worse, a struggle with data consistency.

"Ultimately, it was really hard to set up those connections to Optimizely to make sure we were looking at the same data." —Website Strategy Lead, BigCommerce

If you need to export data from a testing tool, hand it to a data engineer, and wait for them to "crunch the numbers" just to get a result, you are **losing time and momentum**. As the Senior Software Engineer at Zwift noted, their previous process was extremely manual:

- Running an experiment on one platform
- Compiling data
- Finding a data engineer had to jump in for insights

This "Frankenstein" approach creates a culture of hesitation. When stakeholders enter a meeting expecting the data to be wrong or the tool to break, **the vibe goes down**. You can see it in their faces. The doubt and cynicism.

A UNIFIED APPROACH (The Solution)

For these companies (and many more), the solution was to ditch isolated testing tools in favor of **a unified platform** (in this case, Amplitude).

This unified approach makes analytics and experimentation share the same DNA, empowering you to:

- **Solve the identity crisis**
- **Deliver advanced targeting**
- **Power full-stack integration**

Read on to learn how teams made the shift—and what it unlocked.

SOLVE THE IDENTITY CRISIS (The Solution: Part 1)

In mature orgs, a user is rarely just a cookie on a browser. As such, teams need the ability to create holistic customer profiles—aka they need effective identity resolution.

In experimentation, identity resolution requires two key components: gathering accurate and reliable data, and ensuring users experience smooth and consistent experiments.

In essence, identity resolution is the backbone of reliable experimentation.

But most experimentation tools aren't natively integrated with an analytics platform, so they can't provide automatic user ID merging.

At **Huckleberry Labs**, a single account might represent a mother, a father, and a nanny, all accessing the app across multiple devices to track a child's sleep patterns.

Their previous stack struggled with this. It handled assignments at the device, not the user level.

This meant that a mom might see one version of a feature on her phone, while the dad saw another on his, creating a disjointed experience and polluting the test data.

SOLVE THE IDENTITY CRISIS (Continued)

"We needed a platform for A/B testing that had already figured out how to do user-level assignment of the feature flag, not just device-level assignment."
—Director of Business Analytics, Huckleberry Labs

Switching to Amplitude enabled Huckleberry Labs

move from device-level assignment to user-level resolution.

Amplitude Experiment ensures properly attributed results and a consistent user experience, even if they use multiple devices, with no extra engineering lift.

Amplitude uses remote evaluation, so instead of deciding which variant a user will see within the app (e.g., variant A vs. B), the request is sent to Amplitude's servers to make the evaluation decision, tapping into all of your user data in Amplitude.

Now, the Huckleberry Labs' "user" remains consistent regardless of where they log in, ensuring a consistent experience across multiple devices.

ADVANCED TARGETING

VIA JSON PAYLOADS (The Solution: Part 2)

Every experiment needs at least one variant to compare with your control experience. But for Zwift, A/B testing isn't as simple as changing the color of a button. They deal with:

- **C++ game engine** (so no out-of-the-box solutions)
- **Different hardware** (gaming PCs vs. basic tablets)
- **Different client versions** (anonymous users vs. known players)

They needed to ensure that a user on a high-end device saw a high-fidelity experience, while a user on a low-end device saw a performance-optimized version, all without breaking the test groups.

"We rely on the extracted information from the hardware to decide if we want to show this very high-end thing.. or if we show a low-end experience that doesn't break your computer."—**Senior Software Engineer, Zwift**

The solution? **Leverage advanced JSON payloads** within Amplitude. This allowed Zwift to dynamically shape each variant's experience, without writing additional code or shipping new client releases.

Rather than simple on/off feature flags, these JSON configurations define exactly what an experience should look like for a given user or cohort.

For example, during onboarding, Zwift uses a "hero carousel" banner. With JSON payloads in Amplitude, they can dynamically tell the game client:

"If this user has never logged into the Companion App, show them the 'Download App' banner. If they have downloaded it, show them the next logical step."

ADVANCED TARGETING VIA JSON PAYLOADS (Continued)

What makes this possible isn't JSON alone—**it's where those payloads live.**

Because experimentation and analytics are unified in the same platform, targeting logic can reference real user behavior in real time: app usage, device constraints, progression state, and more.

That means cohorts aren't defined by brittle client-side rules or static flags, but by live, trustworthy analytics data.

Advanced targeting only works when **your experimentation platform understands your users.**

And that level of understanding only comes from a unified system where experimentation, identity, and behavioral data all share the same source of truth.

INTEGRATE INTO THE CMS AND PRODUCT

12

(The Solution: Part 3)

For BigCommerce, a SaaS ecommerce platform, running a test used to mean coding it in a clientside editor or handing it off to developers.

Now, they've bypassed this engineering bottleneck by integrating Amplitude Feature Experimentation directly into their visual builder, Makeswift.

A product manager or marketer can go into the visual builder, set up an experiment that communicates directly with the backend, and publish changes without touching code.

It enables teams to test deep backend logic with (almost) the same ease as front-end copy changes.

"I can go into the visual builder... set up an experiment... run that experiment. And once that experiment is done... I can then just go in and publish those changes... rather than having to hand-code all that stuff." —**Website Strategy Lead, BigCommerce**

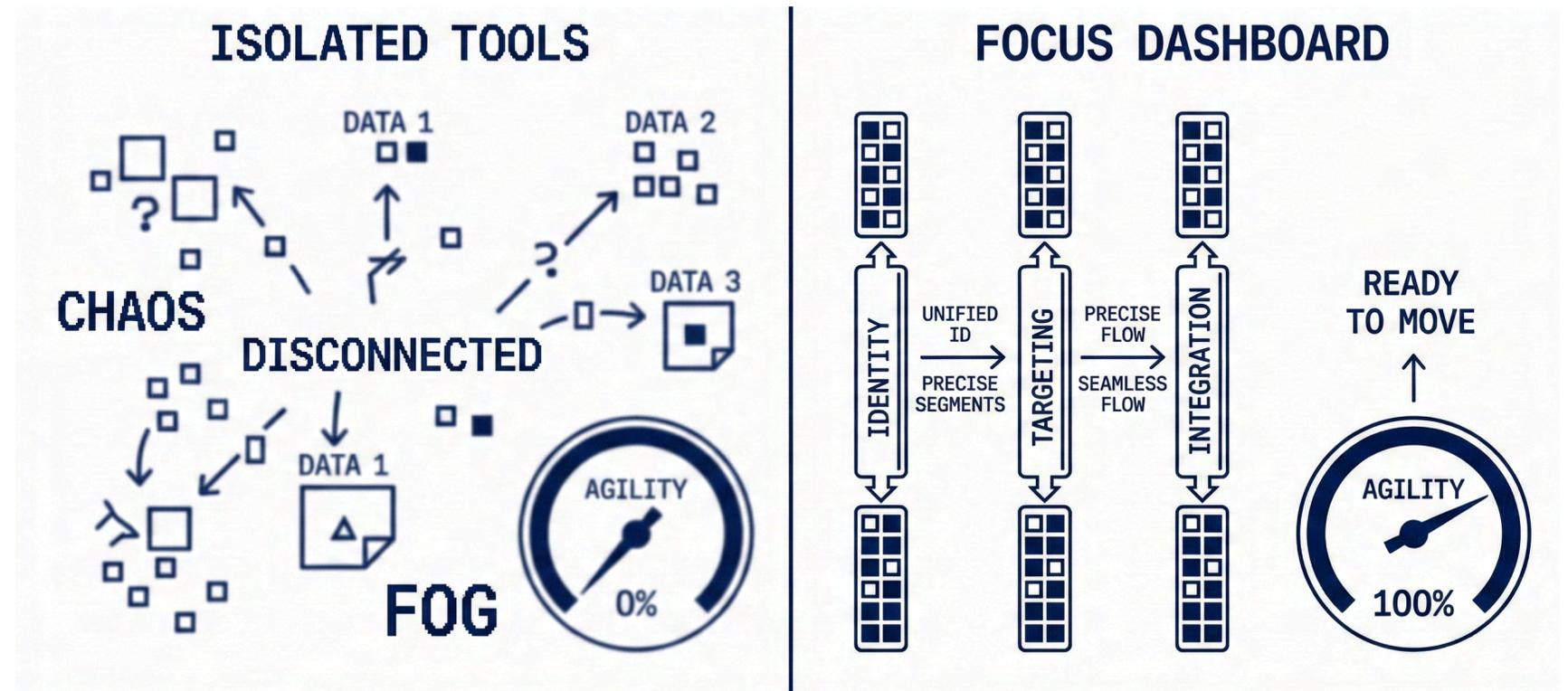
This is the difference between "feature flagging" and true full-stack experimentation.

UNLOCK THE POWER OF UNIFIED EXPERIMENTATION (The Results:)

Resolving experimentation issues related to identity resolution, targeting, and integration isn't just good for product teams; it's good for business.

Adopting a unified approach gives organizations the agility and accuracy they need to make big bets that pay off.

Let's explore how our interviewees unlocked new levels of velocity and trust.



THE VELOCITY UNLOCK

The most immediate impact of switching to a unified approach is **speed**. FanDuel, a giant in the sports betting space, saw its experiment velocity **more than double** in the first year after unifying its stack through Amplitude.

Why? Because the friction of context switching was gone. They didn't have to send engineers to one tool to implement flags and product managers to another to analyze results.

Everything happened **in one ecosystem**. This allowed them to shift their focus from simple "validation" tests (did we break it?) to true "optimization" and "exploration" tests (how do we make it better?).

"In year one, I would say we just over doubled our experiment velocity." —**Product Analytics and Experimentation Lead, FanDuel**

Amplitude's intuitive user experience also enables companies like FanDuel to **unlock self-service experimentation at scale**.

Product and marketing teams can run experiments directly in the same platform they already use for insights, enabling them to optimize faster, personalize more effectively, and ultimately, drive growth.

Meanwhile, your engineering teams don't have to think about one-off requests and can focus on higher-value work.

THE TRUST UNLOCK

Trust will make or break the success of your experimentation program.

If people **doubt experiment data**, they'll default to gut instincts, hesitate to roll out winning variants, or re-run results until they find the story they like.

A **unified platform reinforces trust**, providing one source of truth, data visibility, built-in guardrails, and fewer failure points.

At Zwift, the trust in the new system is so high that they have moved to testing their indoor biking MMO in production.

"We can actually test everything in production leakfree at no risk using a combination of feature flags and configurations." —**Senior Software Engineer, Zwift**

They now have a specific "Employees" cohort. Engineers and product managers can hop on their bikes, log into the live game, and test new features that are live in the production environment but hidden from the public.

This enables them to test the actual end-user experience, find bugs, and fix them before a full rollout.

Why? Because they trust the targeting mechanisms to keep those experiences contained.

Amplitude Experiment empowers teams like Zwift with robust capabilities to test rigorously without sacrificing control or performance. Teams can:

- Roll out & roll back fast with enterprise-grade feature flags
- Deploy client- or server-side, evaluate locally or remotely
- Use built-in support for sequential testing, T-tests, multi-armed bandits, CUPED, mutual exclusion groups, holdouts, and more

THE INSIGHTS UNLOCK

Perhaps the most telling metric of success is how **teams feel about the data.**

At Huckleberry Labs, the mood shifted from dread to excitement.

In one instance, a leader was so excited about a test result—an 18% improvement—*“that he was shouting it from the rooftop of a cabin during a team retreat.”*

When the data is **trustworthy and accessible,** looking at experiment results becomes a habit teams look forward to.

“It gets to a point where checking experiments becomes a highlight in your daily routine. Every time you open your computer, you go straight to the experiment space to see how it’s performing.”—**Senior Software Engineer, Zswift**

This is the aha moment for an experimentation culture. It shifts from a compliance exercise (“Did we check the box on testing?”) to a genuine hunger for learning.

A loss is no longer a failure. It’s a **“discovered revenue”** opportunity, a signal that you touched a lever users care about.

OPTIMIZE THE SYSTEM, NOT THE TOOL

If your experimentation program is stagnant, stop blaming your ideation process. Stop blaming your stakeholders for being "opinion-driven." Stop blaming your individual tools.

Look at your foundation. The teams at **Huckleberry, Zift, FanDuel, and BigCommerce** switched their approach, not just their tools.

Leading experimentation teams measure and deliver user experiences in the same place.

As you look at your roadmap for the coming year, ask yourself:

Is my research system designed to fuel action, or is it just a graveyard for disconnected data?

If it's the latter, the solution lies in your stack.

"Experimentation is inherently a multi-player activity. You want to reduce the friction and bottlenecks that can arise from this as much as possible so you can accelerate your learning,"

says Director of Product Management for Experimentation Eric Metelka.

"With a unified foundation you can do this. You can go from insight to experiment setup to analysis with trusted metrics all on the same platform without additional instrumentation."

POWER EXPERIMENTATION WITH AMPLITUDE

Ready to unlock innovation like these teams did?

Amplitude's AI Analytics Platform is the unique solution for putting products and experiences to the test, learning what your users actually want, and taking action:

Feature Experimentation: Self-service testing that's as simple to set up as it is rigorous and precise in its capabilities, from feature management to T-tests to CUPED and more.

The full Amplitude Platform: Consolidate data models, analysis, and testing workflows as teams use the same trusted data to target experiments and track behavior.

AI Agents: Sense, decide, and act faster than ever before with agents that spot anomalies, propose test design, and roll out experiments for you.

Whether you're starting from scratch or scaling your program, Amplitude gives you **the velocity and trust** to take experimentation to the next level.

[Try Amplitude for free today](#)