Extracted from:

Practical A/B Testing

Creating Experimentation-Driven Products

This PDF file contains pages extracted from *Practical A/B Testing*, published by the Pragmatic Bookshelf. For more information or to purchase a paperback or PDF copy, please visit http://www.pragprog.com.

Note: This extract contains some colored text (particularly in code listing). This is available only in online versions of the books. The printed versions are black and white. Pagination might vary between the online and printed versions; the content is otherwise identical.

Practical A/B Testing

Creating Experimentation-Driven Products

Leemay Nassery

Edited by Vanya Wryter

Practical A/B Testing

Creating Experimentation-Driven Products

Leemay Nassery

Pragmatic Bookshelf

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and The Pragmatic Programmers, LLC was aware of a trademark claim, the designations have been printed in initial capital letters or in all capitals. The Pragmatic Starter Kit, The Pragmatic Programmer, Pragmatic Programming, Pragmatic Bookshelf, PragProg and the linking *g* device are trademarks of The Pragmatic Programmers, LLC.

Every precaution was taken in the preparation of this book. However, the publisher assumes no responsibility for errors or omissions, or for damages that may result from the use of information (including program listings) contained herein.

For our complete catalog of hands-on, practical, and Pragmatic content for software developers, please visit *https://pragprog.com*.

The team that produced this book includes:

CEO: Dave Rankin
COO: Janet Furlow
Managing Editor: Tammy Coron
Development Editor: Vanya Wryter
Copy Editor: L. Sakhi MacMillan
Layout: Gilson Graphics
Founders: Andy Hunt and Dave Thomas

For sales, volume licensing, and support, please contact *support@pragprog.com*.

For international rights, please contact *rights@pragprog.com*.

Validating Longer-Term Impact with Holdbacks

Usually, we want to retain results as quickly as possible. However, sometimes we need more time to demonstrate the longer-term impact of changes on business and product metrics. Or maybe we want to quantify the collective impact of all product changes made within a quarter. There are also scenarios where we want to ensure the initial results continue to trend in the same direction over a more extended period. For the insights that may not yield results as quickly, this is where long-term experiments come into play.

By opting for holdbacks and long-term experiments, you can accomplish the following:

- Measure the impact of changes on metrics, such as churn or retention, that alter at a slower rate or take longer to observe.

- Learn the relationship between your short-term product metrics and long-term business metrics in a casual manner.

- Understand the impact of multiple changes collectively, given that the holdback group has yet to be exposed to them.

- Continue to observe that the trend from the initial A/B test either maintains, improves, or degrades given the initial results.

Let's first better understand what holdback experiments entail.

Defining Degradation Holdbacks

If you want to evaluate a feature's impact on a longer timeline that has already been launched to most users, then you should select the degradation holdback, also called holdback.

For a degradation holdback, most users will receive the feature. At the same time, the feature is unavailable for a small percentage, the holdback group. Reasons to run a degradation holdback include the following:

- If you're comfortable removing the feature from the experience for the set of users in the holdback group and potentially degrading the experience if that feature is key to the product experience for a longer time period.

- If you require the flexibility to select which combination of features to remove for your evaluation.

- If you want to know the impact of a feature on certain key business metrics that take longer to observe, such as three to six months.

Using a Degradation Holdback for the For You Homepage

To further illustrate when to use degradation holdbacks, let's return to the For You homepage at CableMax.

It's unclear what the long-term impact is of the For You homepage on the video product. To attain this insight, the For You homepage would be the perfect candidate for a degradation holdback.

The For You homepage was a significant change to the CableMax video product. It's important to use a degradation holdback to mitigate the risk of unintended effects by monitoring key metrics such as long-term retention and video consumption. By employing a degradation holdback, you'll be able to answer the following questions:

- Will the initially observed increase in consumption persist over time?

- Will other user cohorts benefit from the For You homepage that may have taken more time to become familiar with the feature?

- Will the For You homepage influence retention, whether positive or negative?

Knowing how the For You homepage affects metrics that take longer to observe, such as retention and churn, can also influence future changes to the video product. Suppose the results of the long-term experiment suggest that the For You homepage has a positive influence on user retention. In that case, the team could further prioritize similar innovations to improve user retention. It's a great example of when to use a degradation holdback, leading us to define when not to use this type of long-term experiment.

Clarifying When Not to Use a Degradation Holdback Test

You should not choose a degradation holdback A/B test if your feature is already available for your entire user base and removing the feature for existing users to create the holdback group results in disorientation with the product experience. This could be the case for features or changes that are highly visible on the product.

For example, think of common websites you visit on a daily basis and pick one to use for this exercise. For instance, websites that you might visit daily could be google.com or youtube.com.

Now think of removing the homepage or landing page from the website for A/B testing purposes. How would people behave or interact with the site if the homepage were missing? What would you do if you visited the site and realized the homepage was unavailable? Suppose the homepage were missing

from google.com. In that case, some people may think their computer's mal-ware protection software failed, causing a virus to seep its way through their browser history, one website at a time, modifying the CSS rendered. Or maybe they would restart their computer, over and over again, until google.com was restored to its original homepage.

The effect of removing a prominent feature would likely outweigh any insights you could gather if you launched a degradation holdback A/B test—unless there was a workaround that would alleviate this concern, which you'll soon see in the next section.

Leveraging the Power of New Users

A super-secret workaround if you want to implement a degradation holdback test for a highly visible feature (not so secret now) is to alter the eligibility criteria for the holdback user group to specify only *new users*. This would avoid the disorientation effect, as new users have yet to engage with the feature that's removed.

Another benefit of incorporating new users into the test and control variants is combating the Hawthorne effect or "newness" bias. Let's say a new feature is introduced, and longer tenure or existing users' interests are piqued. So, they click everywhere, over-indexing time spent initially on the new feature. This wouldn't be the case for new users who are unaware that a feature is new. New users are more likely to be less sensitive to product changes than existing users because existing users have created habits because they're more familiar with the product.