3-Part Whitepaper Series: The Top Objections to Testing (and Why They're Wrong!)
By Brian Tvenstrup

Part 2: Objections - Testing is Expensive and Uncertain!

Now that we have introduced the idea of testing and considered the most fundamental objection to its use, we will consider a set of other objections that are commonly utilized when arguing against the need for testing. These objections largely revolve around the two perceived pitfalls of expense and uncertainty. Like many objections, there is often a grain of truth at the core, but through a more detailed investigation of these claims, we will consider whether they truly constitute fatal flaws against testing.

Objection #2. If I can't control for everything, then testing is worthless.

This is a very common fallacy, but at its core, it is mainly an artifact of differences in language usage between professional statisticians and others. When analysts or statisticians talk about controlling for outside factors (sometimes called "exogenous factors"), they do not necessarily mean that that everyone in the test groups is being treated exactly the same with respect to everything else except for the specific treatment being tested. This is impossible to do, and if it really were required to generate valid results, then testing would have limited value indeed! What is meant by controlling for other factors is that whatever those factors are, they have comparable or equivalent impact on the groups being compared within a test. For example, you could be running multiple promotions for a website, offering some high-value customers free shipping on their orders, while others do not receive the offer. Then you decide to test sending out an additional email to highlight a new product. So you identify the group of customers to whom you want to send this additional email, and then divide that population into two groups, one which will receive the new email, and one which will not. As long as your two test groups have been selected properly (primarily through a randomized process), then they will have a nearly equal proportion of customers who are also receiving the free shipping offer. Consequently, when you end up comparing the group that received the new product email to the one that didn't, then the impact of the free shipping offer will be comparable between the two groups and won't bias the analysis in either direction. (Of course, if you've been listening to the advice in this whitepaper, you would also be testing the free shipping offer rather than just offering it to everyone you think should get it!) This same logic holds for any combination of outside factors: as long as the impact is comparable across test groups, then they should not affect the ability to get good results from the test in question.

An important corollary to this objection is that you do need to structure the tests in a way that allows the influence of these external factors to be minimized. Mainly, this is a function of two variables: the size of the test and the length of the test. Tests need to be large enough that idiosyncratic differences between groups do not show up as differences which you would attribute to the test itself. All tests should be designed to provide results from which meaningful

conclusions can be drawn---which is the same as saying that the results should lead to outcomes that are better than guessing! To do this properly requires analysis of issues related to sample size, variance, statistical power, and probability values or confidence intervals. Since this is not a technical paper, we are not going to go into the details of how these factors are calculated or the tradeoffs between them. Suffice it to say that it is critical to consult with trained statisticians when designing tests to ensure that they are optimized to allow these sorts of meaningful conclusions to be drawn. Even the most gifted statistician in the world can't repair the damage if only given results after the fact and asked to draw conclusions, if the original test design was flawed.

Tests also need to run long enough to generate robust data. This is about more than just accumulating enough counts; it is also about allowing the influence of exogenous factors to be minimized. The shorter the time period of the test, the more likely that some aberration in an external factor can affect results. Random effects of things like production delays, seasonality of demand, other market shifts, and even the weather, all can influence a test over the short term. But over time the impact of these factors tends to get washed out, so running the test for an adequate time period helps minimize potential bias introduced by random effects.

Objection #3. Business conditions change too fast to make testing worthwhile.

This is another objection that is more theoretical than practical. Some have tried to caricature making business decisions based on test results by comparing it to driving by looking in the rear-view mirror. While witty, this analogy is fundamentally flawed, because the two alternatives of testing versus not testing are not comparable to driving while looking backward versus driving while looking forward. Instead, the analogy would be more accurate if testing was compared to driving with a rear-view mirror, while not testing was compared to driving with no mirrors and no front windshield either! By now it should be clear that testing yields valuable information for the business that is available through no other means. Of course, having this information does not equate to possessing a crystal ball that allows business decisions to be made with perfect foreknowledge. But given the uncertainty about the future, any decisions that need to be made will be better informed by the wider range of data that testing provides compared to relying only on BAU results.

Moreover, testing is a continual process, not a "once-and-done" effort. Technology, costs, pricing, competition, and other factors are all continually changing, so what worked three years ago (or even three months ago) may no longer be optimal in today's business environment. Consequently, the more rapidly the business environment is changing, the more likely that BAU strategies will devolve into sub-optimal treatments, even if they were perfectly set in the first place. But without testing, it would be impossible to know how much and how rapidly existing strategies are deteriorating. A good program of continual background testing will enable adjustments to be made on the best available information at any point in time, and "what-if" analysis can be informed by knowledge of the prior empirical effects of the variation of key

tactical elements. Testing will also provide more concrete data on which to base forecasts of future developments and make adjustments accordingly. So the more you fear that your business conditions are subject to change, then the more valuable testing will be for your business.

Objection #4. I already know what works for my business, so I don't need to spend time or resources testing to give me an answer I already have.

This objection is often rooted in a view of the business world that pits testing, analytics, and empirical analysis against business judgment, intuition, and experience. But properly understood, testing is not a replacement for business judgment---it is a tool to help improve business judgment! In fact, the more testing that you are able to do, the better you are able to incorporate the learning generated into your future perspective when similar issues arise.

Testing also represents a way of "accelerating" experience. Some businesses set their tactical plans once per year, based on their current strategy and view of the marketplace. They execute against that strategy all year long, and then decide at the end of the year what changes to make for the subsequent year based on their perspective on what worked well and what didn't. So over five years, you might be exposed to five different tactical plans and develop a perspective on what worked and what didn't. But imagine if your business executed versions of all five tactical plans in the first year! After only one year, you would have gained as much experience with those tactics as you otherwise would have taken five years to accumulate. Imagine how much faster you would learn, improving your business judgment for the future, based on this acceleration of experience.

Intuition is really just the mind's subconscious decision-making process that occurs below the level of conscious reasoning---but its accuracy is still largely a function of prior experience and knowledge. Rarely does intuition produce flashes of genius in areas with which you have no prior experience or knowledge. So by broadening your perspective in a shorter time period, testing will actually hone your intuitive accuracy over time as well.

Of course, we would be remiss if we failed to consider the simple fact that, sometimes, business judgment or intuition can be wrong. Without a framework for testing, it can be expensive to discover this is the case, or it might take a long time for the error to manifest itself. In extreme cases, certain types of errors (mostly those of the sub-optimization type) may never be discovered at all without the benefits of testing. So testing performs an important corrective and validation function for business judgment as well.

Objection #5. Testing is too expensive for my business strategy.

Of all the objections to testing, this one has the largest grain of truth at its core. Many executives understand all the conceptual groundwork that has been detailed during the discussion of the previous objections. Yet they resist testing because they believe that the costs of testing are too

high for their business. But in actuality this is a classic case of a "penny-wise-and-pound-foolish" decision.

In the case of testing, the costs of testing are typically known in advance. The cost of testing includes both its direct costs (e.g., how many pieces need to be produced) as well as its opportunity costs (e.g., what is the cost of the discount being offered compared to the standard price). Estimating these costs is not usually very difficult, and the number will usually be positive. So an executive has to accept a known cost of some magnitude to derive an unknown benefit, and some people, in this circumstance, find it difficult to accept the cost of the test.

However, this is an asymmetric view of the situation which unfairly burdens the test simply because its costs are more visible to the organization in advance. The basic problem is that the errors in BAU (those associated with sub-optimal decisions) could be high, but without testing they remain hidden, so they are not weighted appropriately against the known costs of testing in advance. As discussed previously, only testing can expose the true costs associated with sub-optimal decisions that would otherwise be made, so to some extent, not testing is a way of keeping the organization in ignorance of these costs, so they can't be weighed against the known costs of testing. A vicious circle can ensue, in which every proposed test can be shown to have associated costs, while any suggested benefit from the results of the test can be discounted due to uncertainty, such that the organization never actually implements any tests at all. Of course, in some instances, a test tactic may be suggested which actually has lower direct costs than BAU, and tests such as these ordinarily have a higher likelihood of being pursued. However, it is not necessarily better for the organization to pursue tests that have lower direct costs than their then-current BAU practices more aggressively than other tests, so wise executives should be aware of this bias.

Ultimately, good testing as a whole is a very highly positive ROI activity in almost any organization, because the value of the information gained through testing allows decisions to be improved and the resulting differences in the business outcomes are far greater in magnitude than the associated costs of implementing the tests. However, not every individual test necessarily passes the same return threshold because not every test will yield information that has a significant impact on the business results. So business leaders must avoid the temptation to ask each test to cost-justify its existence prospectively. Of course, every test should have at least a plausible outcome that would lead to improvements for the business, but if analysts knew in advance what the tests would show and which tests would be successful, they wouldn't need to perform them in the first place! And the more complex the existing BAU tactics, the more likely that some kind of testing is needed to determine how they could be improved by alternative treatments.

The best organizations understand this framework and simply factor the cost of ongoing testing into their assumptions about the cost of doing business. The reality is that good tests are very efficient---even a few hundred observations can provide solid directional learning, and if you can afford alternative treatments for thousands of potential targets, then results should be very

informative from a statistical confidence perspective. Well-designed and robust test plans can operate very effectively over time without consuming more than a few percentage points of the budgets of their relevant business units, while yielding benefits many times over.

Now that we have reviewed the most common objections to testing, in the final installment in this series, we will review considerations for the successful implementation of testing in a typical business environment.