

Tracking Search Epiphenomena: Why This Behavioral Anomaly is Vital in Web Analytics

Written on October 13th 2006

Author by [Gary Angel](#)

I'm pretty sure that with this title I'm going to skyrocket to the number one position on Google – for “epiphenomena.” Laugh while you can, because epiphenomena are actually vitally important for understanding web analytics – especially if you are an ad-based web site or, indeed, any website that doesn't have clear and obvious conversions.

In my last installment, I described a host of problems that beset sites when trying to use conversion as THE metric of choice. But no matter how many issues surround tracking conversion, it's still a lot easier to deal with and more powerful than anything available to sites that don't have conversions as a convenient yardstick for success.

For ad-based sites, there is no such thing as a single conversion event. When a page is a billboard (banner-based), then the publisher most often makes money on a per-impression basis. In this model, page views are the primary driver of revenue. But there is a secondary dynamic – the more visitors click through on the banner, the more likely advertisers are to be happy. Yet every visitor that clicks through on a banner is lost to the site – so no further impressions get recorded (in that session). In other words, there is a built-in conflict in the basic site goals – the publisher benefits most from impressions without click-throughs but the advertiser benefits most from click-throughs.

For content-network ads placed on a pay-per-click basis, the publisher only makes money on the click-out. That's straightforward enough, but if a site mixes banners and PPC links, then PPC links have the same impression loss that banners do.

The more engaging content is and the more tantalizing navigational paths that a publisher provides, the more page views (and impressions) a site will drive. But engaging content and tantalizing navigational paths may reduce click-outs – making advertising less successful and therefore less profitable.

So unlike ecommerce sites, ad-based sites have to struggle with a conversion event that is amorphous (revenue is driven by every page view) and contains within it at least two contradictory impulses (maximize views vs. maximize click-throughs).

For the most part, publishers resolve these conflicts sensibly enough by focusing on driving as much consumption of content as possible while selling to appropriate advertisers and letting “nature” take its course.

There is, of course, a usability tradeoff when it comes to maximizing impressions. The quicker and faster a visitor finds the content they are looking for, the better for them and the higher their likelihood of return. But the quicker and faster they arrive at a destination, the less impression opportunities they provide.

Most ad-based publishers recognize that optimizing the site experience – even at the expense of short-run impression maximization – is almost always the right strategy.

But here is where epiphenomena come in. Because in trying to figure out how valuable content areas are, publishers (or their web analysts) often completely miss the boat.

So just what are epiphenomena? Epiphenomena are effects or symptoms which are associated with a deeper effect or condition, but are not the primary cause of this condition. For example, rich people drive more BMW's than poor people. But driving a BMW isn't what makes you rich (quite the contrary)! Catholics are more likely to have a Christmas tree in their living room during the winter holidays. This doesn't mean that setting up a Christmas tree in somebody's living room will make them a Catholic.

On a website, epiphenomenal effects are incredibly common and surprisingly easy to misunderstand. In one recent analysis, we at SEMphonic segmented visitors based on whether or not they viewed some specific content areas on the site. The goal was to understand how effective these site areas were in driving same session impressions, overtime visits and lifetime impressions.

For one content area we tested, the average visitor had seven lifetime visits and averaged more than sixteen pages per visit. Both numbers are way above the site average. Is this content area a winner? The question is a trick one – because I haven't given you any of the essential information to answer it.

Why not? Well, it turns out that in most cases viewing a content area is entirely epiphenomenal to the kind of statistics I'm talking about here. To get to the content area in question, visitors tended to be dramatically more engaged with the site. In other words, the visitors who found the content were the sort that were already quite likely to have many visits and view lots of pages. The content area view didn't cause the type of visitor – the type of visitor caused the content area view.

In fact, by implying that we found one content area like this I've misrepresented the case. Because we found many, many content areas like this - which is part of what helped convince us that the effect was epiphenomenal and not causal. For ad-based sites, smaller, less frequently visited content areas will almost always look like big winners when viewed on an average visit/view basis.

So how can ad-based sites understand when an effect is merely epiphenomenal and not central? We've found several methods of analysis to be much more fruitful.

First, it's important to understand that a content area might be successful in more than one way. Some content is inherently single-use - once a visitor has consumed it, there is no reason to expect them to come back. For single-use content, the type of success we'd expect to see is measured most effectively by subsequent visitor returns without viewing the target content.

On the other hand, many kinds of content area are only working well when they are gathering repeat viewers. For this kind of content, the type of success we'd expect to measure is subsequent visitor returns to the original area.

Let's tackle this second metric first. In a sense, it's a measure of the mindshare that a particular content area grabs from total site visits. And to measure it, we track how many of a visitor's visits (and pages) include the target area. By measuring "mindshare" over

time, we were able to screen off the vast majority of epiphenomenal effects – and establish a much better measure of how influential a content area actually is on return behavior.

But what about tools (and content) that are inherently single use? For this type of content, we've found that the most effective analysis is essentially to construct an artificial A/B test. We construct a population segment of first-time content area users and then create a matching set (across key variables like pages viewed and prior sessions) of visitors who didn't see the content. By measuring the difference in subsequent over-time return behavior, we have a measure of whether or not the content actually improved (or degraded) visitor loyalty.

It's unfortunate, but the vast majority of visitor segments and web KPIs are extremely vulnerable to measuring effects that are mere artifacts of the site navigation and the types of visitors that are most likely to stumble on them. This problem is exacerbated by the restrictions that many web analytic tools place on the complexity of the rules that can be used to create a visitor segment and in their limited ability to track the sequence of events over-time.

For ad-based sites, segments that are built off simple rules about viewing content and use common core KPIs like average visits or page views are almost guaranteed to be misleading. Segmentation (like conversion) has often been put forth as the "right" and necessary way to do web analytics. And, of course, it is very important. But just as measuring versus conversion is far more complicated than it might at first seem, so too can segmentation be badly mishandled. Taken simplistically and measured against the wrong type of KPIs, visitor segmentation can steer both analyst and publisher disastrously wrong!

All of which make epiphenomena pretty darn important. So now, I'm going to go check Google and see if I'm ranked ahead of the Wiki...