GARY ANGEL

MEASURING THE DIGITAL WORLD

USING DIGITAL ANALYTICS TO DRIVE BETTER DIGITAL EXPERIENCES

Measuring the Digital World

Using Digital Analytics to Drive Better Digital Experiences

Gary Angel

Publisher: Paul Boger
Editor-in-Chief: Amy Neidlinger
Executive Editor: Jeanne Glasser
Development Editor: Natasha Wolmers
Cover Designer: Alan Clements
Managing Editor: Kristy Hart
Senior Project Editor: Lori Lyons
Copy Editor: Hansing Editorial
Proofreader: Debbie Williams
Senior Indexer: Cheryl Lenser
Senior Compositor: Gloria Schurick
Manufacturing Buyer: Dan Uhrig

Publishing as FT Press
Upper Saddle River, New Jersey 07458

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Printed in the United States of America

First Printing December 2015

ISBN-10: 0-13-419508-6
ISBN-13: 978-0-13-419508-7

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Control Number: 2015951825

We all need a little pushing sometimes.
This book is dedicated to Grace, Isabella, and Ilise
(my wife and daughters) who, knowing I wanted to,
bugged me until I agreed to write a book.
They will probably make fun of my picture on the book jacket,
but given that it is about digital analytics,
none of them will likely ever read it.
Yet without them, it would never have been written.

Contents

Acknowledgments

There are countless people who have contributed to and helped shape my understanding of digital measurement over the course of almost two decades. There are too many to count and probably too many to remember. I can't not mention the small group of people who helped me build Semphonic, the digital analytics practice where nearly all of this thinking grew up and matured. Every person on that team is special to me. They helped build a company and practice that proved to be the central endeavor of my adult life, and almost without exception they have been good friends as well.

Outstanding among that truly wonderful group is, of course, my original partner and co-founder in that endeavor, Joel Hadary. No matter the intellectual interest of a journey, it's the people you travel it with that make it enjoyable. It's hard for me to imagine a better partner for building a company and creating a discipline.

Three early employees who have been with me now on this journey for more than a decade must also be mentioned. Paul Legutko, Phil Kemelor, and Jesse Gross have all contributed mightily not just to the practice we built, but also to the thinking and approaches laid out here. Paul has shaped so much of my thinking on analytics—particularly the problems of segmentation, selection, and bias—that's it hard for me to know where my thinking starts and his ends. To Phil I owe much of the broader picture that makes analytics more than number crunching and focuses it on people and problems and how to solve them. Jesse—who began with us right out of college knowing no more of digital analytics than I know of Sanskrit—has been at least a full partner in the exploration of voice of customer and its role in digital measurement.

The voice may be mine, but the ideas are, as often as not, theirs.

When I started in digital analytics, it was a tiny community. Like most tiny communities, its members were oddballs. The best kind of eccentrics. Lacking credentials and degrees, digital analytics was peopled by those willing to explore a discipline that offered neither safe position nor any clear path to success. That's often the best kind of people. It has been both a pleasure and privilege to grow up in that community. To argue, to fight, to learn, and to share with a group of people who actually love what they do and strive to do it better.

Thank you all.

About the Author

Gary Angel currently leads Ernest & Young's (EY) Digital Analytics Practice. EY acquired Gary's previous company, Semphonic, in 2013. As President and co-founder, Gary led Semphonic's growth from a two-person consultancy to one of the leading digital analytics practices in the Unites States. Voted the most Influential Industry Contributor by the Digital Analytics Association, Gary writes an influential blog (http://semphonic.blogs.com/semangel), has published numerous whitepapers on advanced digital analytics practice, and is a frequent speaker at industry events. Over the last two decades, he's helped create and advance the state-of-the-art in digital measurement.

Preface

Measuring the (Digital) World

How do we understand the world around us? It's not so obvious.

Our senses feed a never-ending stream of data into our brains. In its raw form, that data is incredibly complex. Shapes, surfaces, sizes, shades, motions, scents, textures—we absorb them all, seemingly without effort. Even the most powerful supercomputers yet developed cannot begin to compete in real time. But the plastic mind of human babies—with the help, no doubt, of some careful pre-wiring—can learn how to make sense of this data, parse it, and react swiftly and intelligently to this tidal surge of information.

For centuries, philosophers have understood that the very existence of a world outside ourselves is impossible to verify. We know the world only from the inside, from the endless, constant processing we call consciousness and the apparent flow of data that we believe is generated by our eyes, ears, nose, and fingers. That such a thing as the physical world exists is beyond our ability to seriously doubt. But we know it only by interpreting essentially abstract data.

We grasp the world by imposing patterns on it—foundational patterns that have been imprinted and wired deep into our minds in the endless laboratory of nature. Size, shape, color, and motion are just a few of the core building blocks of our human understanding of the physical world. As fast as sense data funnels into the brain, we are contextualizing it, categorizing it along key dimensions, and then measuring and comparing everything.

That's the physical world. Evolution ensures that we know how to understand it, even if we don't know how we know. But the physical world is no longer the place where we spend all our time. We also live in the digital world, a world with different rules, different types of data, different frames of reference, and different types of measurement.

For the last 17 years, my job has been to measure the digital world. To glean, from the vast streams of data it showers upon us, the fundamental categorizations that matter. To develop the framing devices, dimensions, and measurements that let us understand this digital world with the same ease and power with which we parse the physical world around us. It's important work because we can shape and improve that world only when we understand how it works—not how it works from a programming perspective (although I started out as a programmer), but how it works for people.

These days, we spend immense amounts of time, energy, and money trying to improve the digital world. How well we do that work can determine the success and direction of public policy (healthcare.gov), the health of our love life (eHarmony), and the state of our knowledge about the world (nytimes.com).

This work isn't easy, and it isn't finished. We are still like newborn babes learning to parse the data from our digital sensors.

The digital world is fascinating. And unlike in the physical world, we have no pre-wiring for measuring and understanding the digital realm. Without that huge prebuilt advantage, our interpretations are wrong much more often than they are right. But when we get them right, we at least know what we did.

The pages that follow show you what we've learned so far about how to measure and understand the digital world.

1

Digital Meaning

We have tools dedicated to measuring the digital world. So it's no surprise that we assume the measurements those tools give us are the right ones for the job. They aren't. The standard set of web metrics most digital analytics tools use were developed long before people had even a basic understanding of how to do digital measurement, and mostly before analytics tools that could do much with the data were in widespread use.

The Digital Challenge: Our Metrics and Our Measurement Lack Meaning

The most common digital metrics are almost useless. They measure the wrong things in the wrong ways. They fail, at the most basic level, to link what happens in the digital world to our understanding of people's behavior. In this chapter, you identify the basic challenge of digital measurement and analytics, and you see why common metrics and reports can't easily answer the fundamental questions you're likely to ask about the digital world.

The Grocery Store with Invisible Patrons

Imagine a fairly normal grocery store, well stocked with cereal, milk, beer and wine, eggs, ice cream, canned goods, vegetables, fruits, and, of course, the usual assortment of treats near the register. Now imagine that the patrons and their carts are invisible. You see the door swing open when they arrive. You hear the cash register ring when they depart. You know what they bought. But everything else remains hidden. It would be hard to know how well the store was working and what you could do to make it better. Is it missing items or brands shoppers want? Is the store laid out in a way that makes life easy on customers? Does it maximize their purchasing behavior? Have you allocated the right amount of shelf space to each type of item? What might you do to get an individual customer to spend more or be more loyal?

These are the types of questions that merchandising experts have studied, pondered, and worked on for many years—since well before the digital world ever existed. Interestingly, they found that they could answer some of these questions even when the customers were, for all practical purposes, invisible. Equally interesting, they found that some types of questions are much harder to answer when you don't know who your customers are and that, for many questions, the data might suggest possible answers but rarely provides definitive guidance.

Suppose, for example, that you found that the most purchased items in your store were milk, beer, eggs, and chips. You might be tempted to move all these items together in one place right at the front of the store. That should make it easy for customers to find what they need quickly and efficiently. Is that the way your supermarket is laid out, with the items you buy most right up front?

Chances are, it's almost exactly the opposite. That isn't because you're invisible! Supermarkets work differently for two reasons. We're all deeply cynical consumers, so you probably identified the first reason right away. Groceries aren't set up for your convenience. They often place the things people purchase most at the very back of the store and might even consciously try to locate them far apart. If you've never made an impulse buy at a grocery store, this might seem odd. But if, like me, you've wandered by the dairy aisle and added some ready-to-bake cookies, or you've thrown a bag of chips next to your beer, it's not too hard to see why this setup works. By trying out different store layouts and measuring how much people buy (their average cart), store designers can maximize total sales. Mind you, most grocery stores count on you to make your decision about where to shop based on other factors than how long it takes you to get your items. They know price, selection, and location are more important than convenience. If a new grocery store opened right next door and had the same selection and same prices, a store might well compete on the convenience of layout. But most stores see their layout as a chance to maximize their profits, not your time.

The second reason grocery stores aren't laid out for your convenience is more interesting and more important than good old profit maximizing. Grocery stores have more than one customer. Guess what? They're all different. When grocery merchandisers began to study what people bought (still without knowing who they were—only what was purchased on the same ticket), they found very distinct patterns. Beer and milk might be two of the most commonly purchased items in a grocery store, but they might not often be purchased together. Chips, on the other hand, go pretty well with that beer. And milk buyers are often looking to add cereal or eggs to their cart. So setting up a grocery with the most purchased items all clustered together might not work particularly well or be particularly convenient for anyone.

What's more, even if a particular setup worked well for you today, it might not tomorrow. When merchandisers could only look at the receipts from each shopper, they had no way to tell how much people's habits and shopping patterns varied. That's a huge hole in their understanding. To get around that, grocery stores created loyalty programs so that, in return for discounts, they could tell what you bought every trip. They found that most people don't shop the same way every time they visit the grocery store. Most of us have regular shopping expeditions when we buy everything we need and go up and down every aisle. Store layout might not be a big deal when we're traversing every inch of the store (or, in my case, traversing two or three times as I remember things). But we also have visits when we've just run out of beer or milk—or, heaven forbid, both. We might stop to pick up lunch or to shop for specific recipe ingredients (my flour, bag of chocolate chips, vanilla extract, and egg visits). These are very distinct types of visits, and it would be great if the grocery store could make each type of visit perfect (or perfect for the grocer). Stores would love to be able to do that. But it's hard to push those shelves around when you walk in the door.

Let's not forget those chocolate bars and women's magazines perched right at the register. Very few of us go to the grocery store with the express intent of buying a Snickers bar and a *Cosmo*, but many of us are tempted by one or the other. What spot in the grocery store do people *have* to linger at with nothing to do but be tempted? That's where the candy (eye and stomach) goes.

We can learn a lot from those grocery store merchandisers when we start to think about the digital world. The straightest path isn't always the best. The customer's goals and your goals aren't always identical. Not every product is the same, and some products are more position sensitive than others. A store doesn't have one ideal layout because it doesn't have one type of customer, and customers aren't always going to do the same thing anyway. Last, and most important, what people actually do tells us a great deal about who they are and why they are doing those behaviors.

Getting Digital

We are blind to the digital world. Unassisted, we have no way of knowing whether our website is thronged with visitors or as empty as a mall after closing time, whether our cash registers are overflowing or stubbornly silent, whether our customers are young or old, whether our content is read with rapt attention or is barely and desultorily skimmed. We need eyes and ears to help us see into the digital world. Certain tools have that very function—to track and make visible the otherwise unseen patterns of that world. These digital analytics tools are powerful and rich. They include hundreds or thousands of possible reports and options that seem to expose every aspect of digital behavior. It's all too easy to forget how dependent we are on the exact nature of those tools and to assume that what they show us and the way they show it to us is all there is.

Our natural senses in the physical world have evolved to give us many advantages. We have adopted a deep and abiding faith in what we *see*. Yet even with our physical senses, it's all too easy to forget that the window they provide into the world is a narrow one.

Remember the image of a dress that went viral in early spring 2015? Many people see the dress as black and blue. Others see it as white and gold. If you look long enough or over some period of time, you might see it each way. If you didn't hear about the dress and you can't believe that anyone could see it differently than whichever way happens to strike you, check it out online and show it around. You'll be surprised.

Optical illusions are just one aspect of how our eyes can mislead us. We see color (no matter how much we might disagree about it), but we don't see heat.

Why should we see heat?

Well, why shouldn't we?

Infrared cameras see heat. It's just another wavelength, and for many purposes, seeing heat is far more useful than seeing light (when hunting at night, for example). For that matter, what if we could see radio waves? Hearing and vision seem fundamentally different to us, but each is a set of waves that different tools inside our body use.

What would our radio-wave eyes make of a Madonna song? Probably not much.

The simple fact is this: Our reality is constrained by the tools we experience it with.

What does all this have to do with digital analytics? The digital analytics tools we have are our windows into the digital world. We see only what they can track or think is important. We see what page a user requested from a server, but we don't see how long that page took to load. We see what link a user clicked on, but we usually don't see what part of the page that user scrolled to. We (sometimes) see what website the user came from, but we don't (usually) see what website that user went to. These choices make a profound difference in how we think about the digital world and what we tend to value as important there.

What if our tools aren't very good? What if the events they choose to capture or the ways they choose to show them to us give us only a shadowy impression of the real digital world or what's important inside it?

I've been around since the very beginning of digital analytics. I witnessed firsthand and, in some small ways, even helped shape how those digital tools evolved. Having seen their history, I know that the decisions about what to track in the digital world and how to track it were often ad hoc and shallow.

The first digital analytics tools were built to read weblogs. These logs weren't built to understand and measure the digital world. They were built to create a record of what a web server was doing so that IT professionals might be able to track down operational problems (although they were hardly ever used for that, either). These logs recorded IT-focused information about which content file was requested, when exactly it was processed, what IP address requested it, how much content was sent, and whether the request was successful.

Because those were the fields in the logs, those were the fields we used when we first built digital analytics tools. And being clever folk, we interpolated a whole lot from this bare bones little set of fields. We figured out a way to group the records by the device requesting them (which we promptly anthropomorphized into a human *visitor*). By looking at the time between requests, we could group the requests into batches by creating an arbitrary time limit, and we labeled these batches of requests *visits*. Then we could look at what page a visitor looked at first in that batch and we called that an entry page. We could also look at what page was last in the batch and call that an exit page.

It's important to realize just how arbitrary these decisions were. When a visitor first arrives on a website, that website sometimes records which website the visitor came from—this is called the referring site. By saving the referring site for each batch of records (a visit), you can get a sense of which sites are generating traffic to your pages. But here's a peculiarity: By defining an arbitrary time limit of 30 minutes to group records, we created situations in which a visit sometimes had more than one referring site; in other situations, a visit had a referring site that was the last page the visitor viewed on the same website.

For example, imagine that a visitor searches on Google or Bing, finds your website, and views a page. Then that visitor returns to the search engine, does another search, goes to a different site, and links from there to you within 30 minutes of the first request. You'll have a single visit with two referring sites. This might sound far-fetched, but in certain permutations, it's not uncommon. Many sessions will have multiple visits to Google interspersed with views of your pages.

It's even more likely, especially in our tabbed browser world (which came after all these definitions were created—you remember browsers without tabs, right?), that a visitor will view a page or tab elsewhere, spend some time there, and then return to your website and view another page. Same session? According to our tools, if that happens 25 minutes apart, it is. But if it happens 31 minutes apart, it isn't. And if it does happen 31 minutes apart, you'll have a whole new visit with a referring site of your own website!

It would have been perfectly plausible (maybe much more plausible) to decide that a batch of records should be separated by a referring site other than your own domain, regardless of time. But that's not the way some early vendors did it, so the definition stuck and became an artifact of truth.

And if a visit is merely a rough-and-ready and poorly defined artifact, then so are the entry page (the first page in a visit), the exit page (the last page in a visit), the visit time (the time between the first and last requests that are part of a visit), and the referring site (the domain recorded in the first record of the visit as the referrer, the site from which the user came)—all based on the way we defined a visit.

As with the words we read on a page, the numbers we see in a tool tend to take on privileged status in our minds. But if the only problem with standard web metrics were a certain sloppiness of definition, our situation wouldn't be all that bad. How much difference can it make whether a visit is defined by 30 minutes of inactivity or a new referring domain? Hard to say.

By showing how arbitrary these standard metrics are in construction, I hope to lessen their privileged status and make it easier to convince you that, not only are they arbitrary, but they are largely misguided.

Web measurement began with weblogs, whose goal was to measure digital assets. This long-ago bias has persisted through every generation of digital analytics tool. The implicit goal of analytics tools is to measure the website or the app. That's missing, if not the whole point, a big part of it. In the digital world, our goal should be to understand our customers, not our digital assets.

Our tools have improved to do that—probably more than our practice has. Digital analytics tools now deliver significant and interesting segmentation capabilities that enable you to define and track cohorts, segment on fairly complex behaviors, and compare different types of users. They even provide limited capabilities for integrating nondigital data into their reporting.

This is all good, and the technology seems to improve continually. But although the capabilities of the tools have improved, the basic views they provide haven't changed much. How many reports in a digital analytics tool tell you anything about customers? For that matter, how many of the digital reports you distribute in your organization have anything to do with customers? And how much do they really help you understand the digital world?

Close your eyes and picture your website. Imagine people of all sorts moving through it. They stop here or there. They go down certain pathways and ignore others. They look at this or that. They make a purchase or head for the exit. Can you see it?

Now open your web analytics tool. Do the reports help you visualize that scene? Do they help you understand who your customers are (beer and chips, or milk and eggs)? Can you find the different types of visits (going down every aisle, or just ran out of something) and see which are most common? Do they help you picture which customers do which visits most often (beer and chips, just ran out, Friday night)? In other words, do they actually help you understand the digital world or do they just confront you with a wall of numbers that elude meaning, even while seeming entirely plausible?

I first started measuring the digital world back when websites were brand new and people still talked about the World Wide Web. I'd spent the previous few years working with a couple large credit card companies, analyzing the way people use their credit cards to create marketing programs (yeah, sorry about all that crappy mail). We used some pretty fancy analysis techniques to group people together, to understand how they used their credit cards and then to classify them. It was pretty easy and powerful. It didn't take analytic genius to know that the cardholder who routinely dropped four figures at

Neiman Marcus was a different beast from that two-digit shopper at the local Walmart.

When I first got my hands on web behavioral data, I ran the same kind of (neural net) analysis and proudly sold the results. But whereas my credit card segmentations had truly been interesting and useful, my digital segmentations looked like some inverted Egyptian monstrosity (see Figure 1.1).

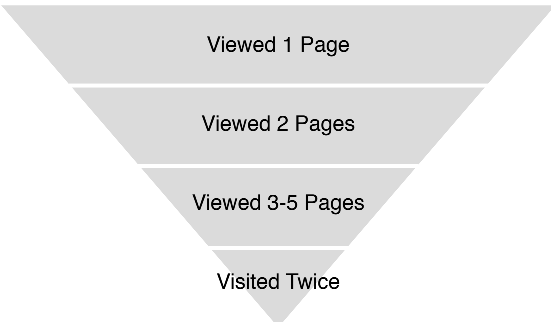

Figure 1.1 Inverted pyramid

Nobody ever got rich targeting "people who viewed 3–5 pages."

I spent years learning that the digital metrics tools provide aren't that interesting and that, no matter how powerful a tool I used to study digital behavior, I wouldn't get interesting results if I picked the wrong type of variables.

So put away your digital analytics tools for a minute. Forget all about page reports, referring sites, average page times, top exit pages, number of visits, average conversion rates, and the whole flavorless cornucopia of web metrics and reports that those tools spit out by default. It's all garbage in the most literal sense—it takes up mental space and it smells bad.

You're about to find a better way to understand the digital world.

Index

D

K-L

M-N

O