



# THE HIGH COST OF CHEAP SAMPLE:

EVALUATING  
THE RELIABILITY  
AND VALIDITY  
OF A PUBLISHER-  
DRIVEN ONLINE  
SAMPLE SOURCE

maru/BLUE

# EVALUATING THE RELIABILITY AND VALIDITY OF A PUBLISHER-DRIVEN ONLINE SAMPLE SOURCE

It has become less and less expensive to conduct a survey, mainly because it has become easier to find people who will answer a survey. But are they the right people? Are they actually representative of the population?

What if the data you gather to help you make a decision is completely wrong? If you consider the cost of a bad decision, what is the real price of that inexpensive sample?

In this whitepaper we look at the reliability and accuracy of a popular survey service that offers the ability to ask Internet users who are trying to access "premium content" to complete a short survey. They must complete the survey before getting access to the content. It sounds reasonable enough: give an incentive and get a response.

**The question is, does this type of sample provide results that are valid and reliable?**

## THE BOTTOM LINE

We found that this sample source was neither accurate nor reliable. Not only were the results wrong, they varied wildly across the five times we asked the same question over the past three years. When we compare the results to the same question tracked by the

Pew Research Center, the findings could hardly be more different.

**This inexpensive sample source is not just bad, it is dangerously misleading.**

# WHAT WE DID

We ran a survey that asked about social media usage and sent it out to a sample of people who wanted access to premium content, at five time periods.

APRIL 2014:  511

DECEMBER 2014:  1,888

JANUARY 2015:  1,510

FEBRUARY 2016:  500

FEBRUARY 2017:  502

In all cases the data was weighted to be representative of the American population, based on their imputed demographics.

The sample was drawn to our study by the sample provider targeting Internet users who seek to access "premium content," including news

articles, videos, or other websites that would otherwise require a payment or subscription to access the content.

The publishers of these websites have agreed to allow the sample provider to administer questions to their users through a corporate agreement wherein the sample provider pays the publisher for access to the potential respondents. In exchange, the respondent gains access to the content for free. The questions appear as prompts when users try to access the premium content; this prompting is also known as a "survey wall" since respondents must either answer the question or click an X to remove the question from their screen. The sample provider uses an algorithm to properly distribute the questions across the publishers' networks.



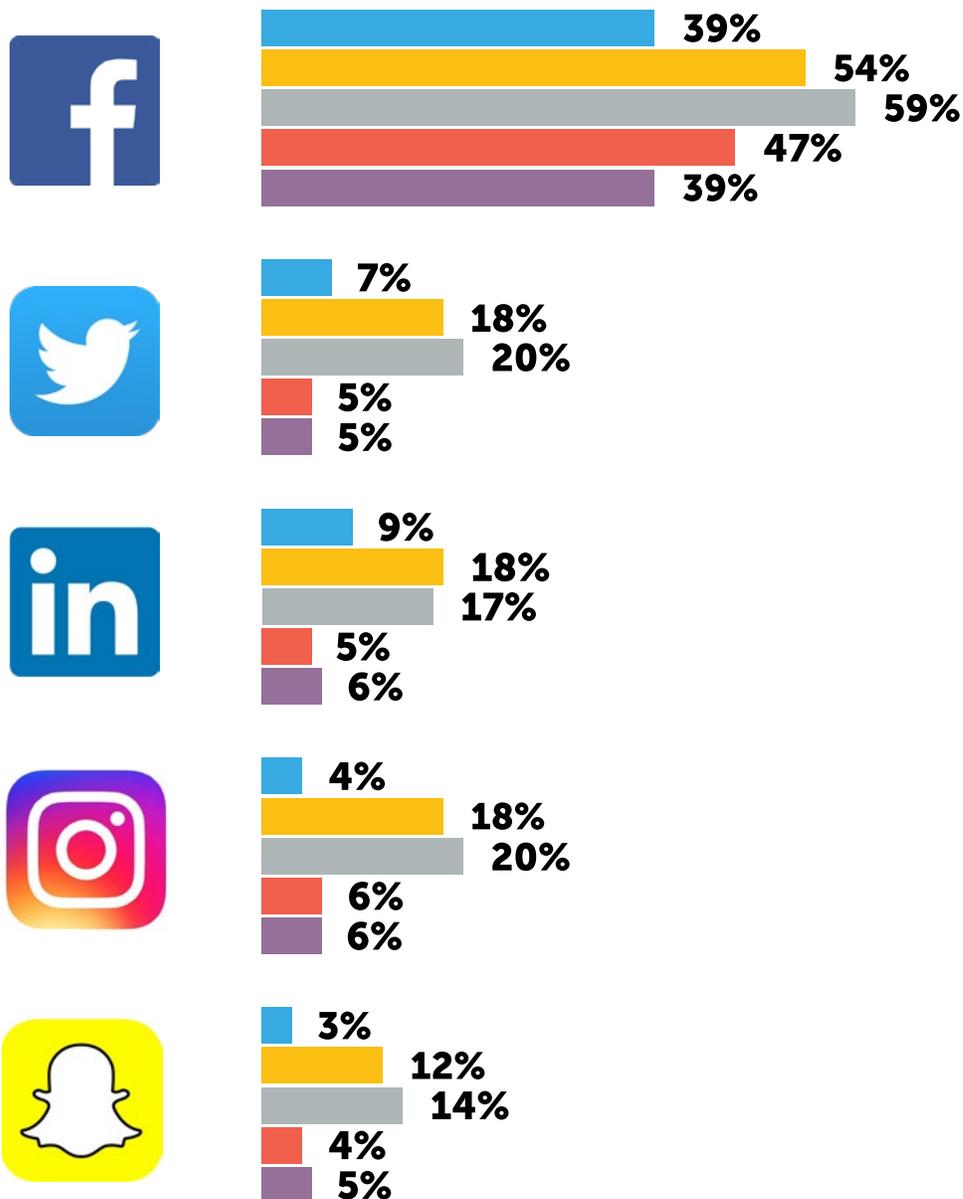
# WHAT WE FOUND

The data suggests that social media grew quickly in 2014, peaked early in 2015 and started into a steep decline in 2016 before falling back to 2014 levels in 2017.

That would be uncomfortable news to a lot of investors and would probably surprise a lot of people who are using these services.

## "Which of the following social networks are you currently active on?" -- sample is publisher source

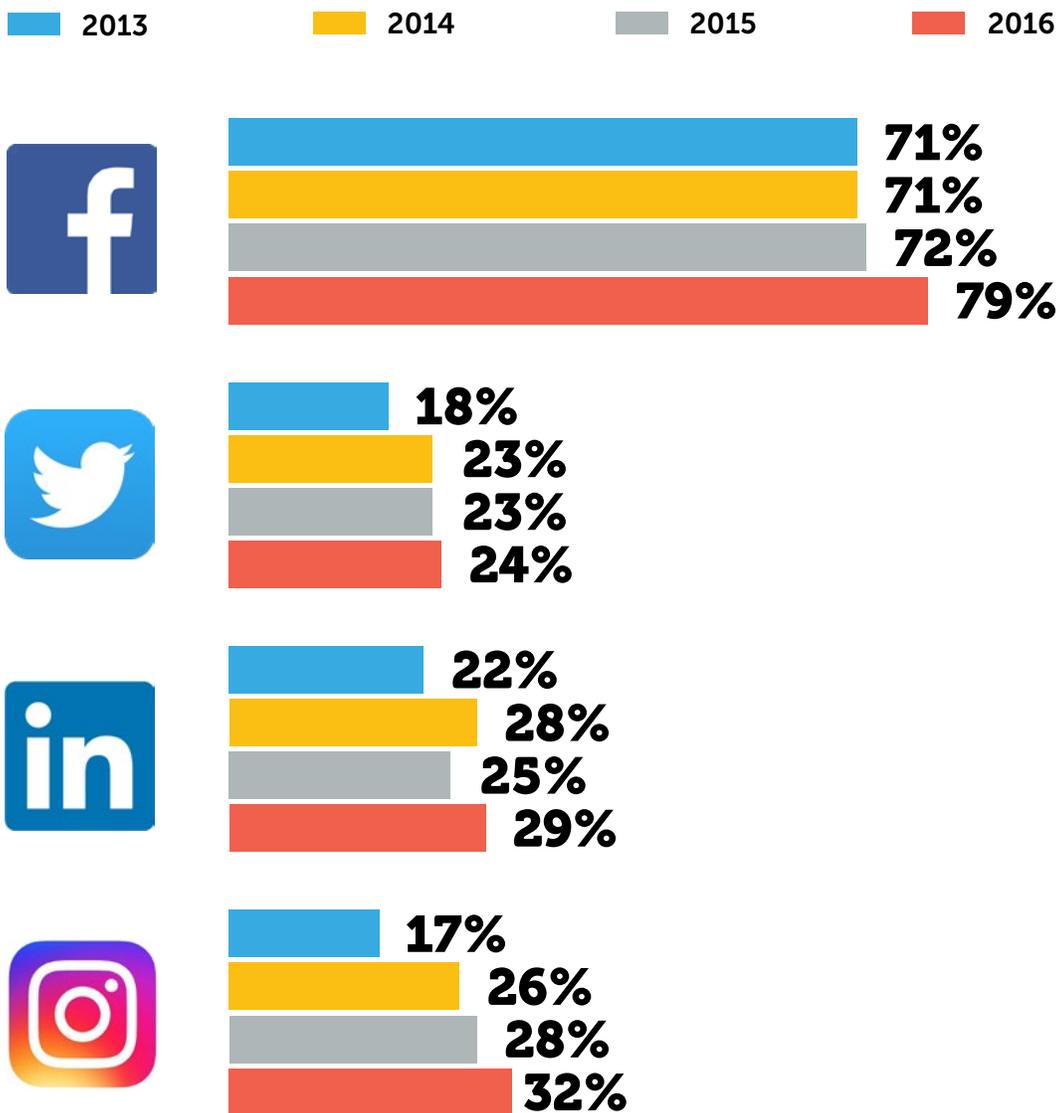
APR-14    NOV-14    JAN-15    FEB-16    FEB-17



# Fortunately the Pew Research Center has been tracking social media use for a few years, giving us a reliable reality check for these numbers.<sup>1</sup>

Their data below paints a very different picture. Not only do they show relatively stable and slow growing levels of usage, their numbers are more than twice as high. They also match what we have measured on our own online community.

## Pew Research Center data on use of Social Media



# WHY ARE THE NUMBERS SO WRONG?

There are many reasons why the data from the publisher sample source could be so wrong.



They could be sampling an unrepresentative set of people.



The data could be weighted incorrectly because of error in the imputation methods used to estimate the demographics.

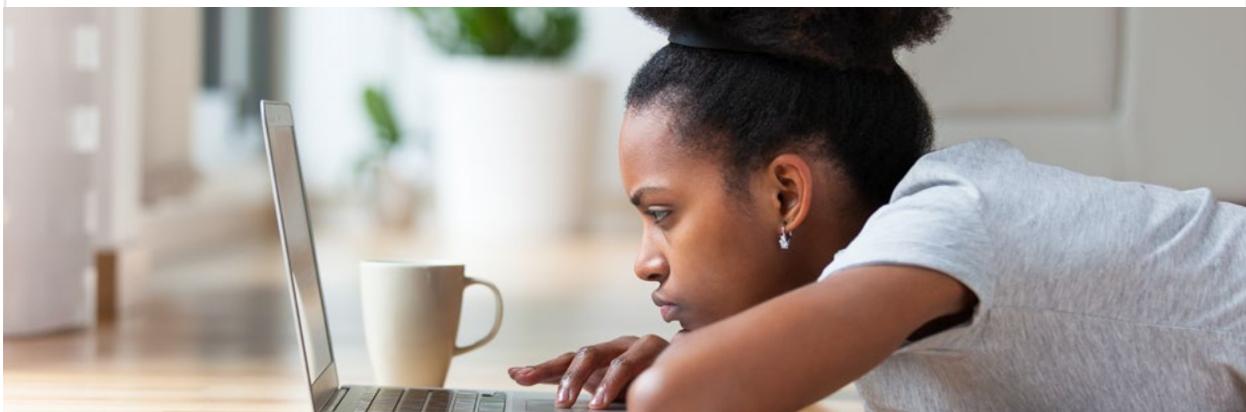
## zzz...

And it could be that the people answering the questions just don't care about the survey, so they don't bother to answer correctly.

**Ultimately, it is probably a combination of all those errors.**

The data might be wrong because the sample might be unrepresentative. The risk of that is high if all respondents came from a single source—like people trying to get access to a piece of content called, for example, “I Hate Social Media.” We know the provider sources from a host of publishers. By sourcing from multiple places it is hoped that kind of bias is at least mixed and muted.

The fact that the demographics are estimated rather than being measured directly is problematic and certainly there are errors. But when we looked at the data weighted and unweighted, the differences were not large enough to account for the kind of variation we observed here. That leads us to the motivations of the people answering the survey.



# WHY THEY RESPOND

We've done a great deal of research on why people respond to surveys, and there is extensive literature on the subject. One school of thought is that people do surveys mainly because they get something material in return—either access or money. Another school of thought is that people respond because they find intrinsic value in making their voice heard and contributing to a community.

Psychologist Anja Göritz conducted a meta-analysis of the effectiveness of incentives entitled "Incentives in Web Studies: Methodological Issues and a Review."<sup>2</sup> She concluded "material incentives increase response and decrease drop-out" but "the combined effect of incentives on response and retention is still small." In fact the impact is very small, accounting for a roughly 3-5% increase in response rate overall, according to Goritz's analysis.

She also reminded readers that "using material incentives is only one option to influence data quality and quantity. We should not forget about other possibly response-enhancing techniques such as personalization, pre-notification, deadlines, reminders, offering result summaries and altruistic appeal." Indeed, personalization, offering summary results and altruistic appeal are large drivers of the intrinsic appeal we see with samples drawn from insight communities.

In our research we found that intrinsic motivations are more important than extrinsic ones—highlighting the value and importance of having an engaged respondent. We found, for example, that:

**89%**  
**AGREE**

I feel like I am doing my part as a good consumer and citizen when I provide feedback

I feel like I am being a trusted advisor when I provide feedback to a company on their products

**87%**  
**AGREE**

**86%**  
**AGREE**

I feel like my opinion makes a difference

I love it when I see the results of a survey I participated in

**86%**  
**AGREE**

**It is all these intrinsic motivations that are fed by being a community member—rather than a person doing a survey just to get money or access to content.**

# WHY SO INACCURATE AND UNRELIABLE?

So why was this publisher-sourced data so inaccurate and unreliable? We know the respondents are not coming from a community that builds their sense of good citizenship, treats them like a valued advisor, gives them feedback and lets them know that their opinion makes a difference.

They are coming for one reason: to get access to content that has nothing to do with the survey they are being forced to answer if they want access.

It stands to reason that if you treat someone like an anonymous commodity, they will respond by treating your survey with the same lack of respect.

**Clearly not all sample is the same. In fact it is not just sample—it's people. And what people want is to be engaged, respected and to know they are making a difference.**



# THINK ABOUT COST IN MORE WAY THAN ONE

When you're look at a sample source that seems remarkably inexpensive, ask about how the sample is sourced and the motivations of the respondents.

And think about the cost of getting incorrect data—information that would guide you to the wrong decision. That's the high cost of cheap sample.



FOR MORE INFORMATION, CONTACT:

[sales@marubluenet](mailto:sales@marubluenet)



<sup>1</sup> [Social Media Update 2016](#)--Facebook usage and engagement is on the rise, while adoption of other platforms holds steady, Shannon Greenwood, Andrew Perrin and Maeve Duggan, Nov 2016

<sup>2</sup> [Incentives in Web Studies](#): Methodological Issues and a Review, Anja S. Göritz, International Journal of Internet Science 2006, 1 (1), 58-70