

## Book Reviews

data, and set the price for using their data. Earning money on valuable data should be normal and easy for the user. The hope is to use and reward social media platforms that promote positive interactions, but he feels that this cannot happen with the current methods of the dominant players, particularly when advertisers are the customers. Consequently, mass deletion of social media accounts is the necessary next step.

Overall, Lanier's arguments are timely for users concerned with privacy, personal choice, and advertisers' influence over their minds and values. Each argument in the book is supported with references, although I would like to see more references to support his eighth argument (although this chapter relies on his previous work, and presumably on the research presented there, it does not stand on its own when compared to other chapters in the present book).

While not written from a Christian perspective, I find it easy to agree with a writer who places people over profits and machines. Lanier presents the costs of using such social media platforms, such as creating people who act less humanely, behave more troll-like, are sadder, and have less empathy for others. The benefits go to advertisers who pay to manipulate users for profit or political benefit. Additionally, Lanier presents arguments that long-term use of social media decreases the user's ability to recognize truth, since platforms remove the context of facts (except for podcasts at the moment).

Consequently, Lanier unwittingly provides an argument that aligns with the Bible's instructions in 2 Corinthians 10:5, where Paul tells his readers to take captive every thought. Similarly, I found that his arguments agree with Romans 12:2, where readers are instructed to renew their minds to test and approve of God's will. It seems that social media have been competing for the attention of our hearts and minds, with purposes at odds with biblical instruction.

Argument ten relates most directly to spirituality and religion, in that he states "SOCIAL MEDIA HATES YOUR SOUL." While not favorable toward religion, this argument highlights parallels between religion and social media. For example, Lanier argues that social media platforms erode users' free will and transfer decision-making power to companies; he compares this to how (he assumes) the church makes decisions for its constituents. Another example is how social media use group mentality thinking to encourage users to treat others badly, similar to how he sees religious conflicts engaging people more

intensely. Lanier sees his suggestions to reject the current version of social media platforms, while not rejecting the core of social media, as similar to the Protestant Reformation during which Protestants rejected harmful practices such as indulgences. Lanier also sees social media as defining truth for its users by selecting the context for facts. He aligns this with the church defining truth within a religious framework, including a reference that some religious people still think the sun orbits the earth. Like religious frameworks, social media provide ultimate purposes for its users, although he mentions that they are poor choices for ultimate purposes, stating that the purposes of Google and Facebook are to organize information and give users a sense of purpose and community.

The part I appreciated from argument ten was the suggestion that people are using social media platforms in a spiritual and religious way. I hope this encourages readers to reflect on the use of such platforms in their lives, so that they can be empowered to use them as a tool, rather than the other way around. As for the church, argument ten observes that social media compete to define truth and purpose for people. This point is helpful as it stands, but the negative portrayal of religion and the church is not. While I realize that the church has issues to work on, it was grating that every issue of comparison presented religion in a negative light without acknowledging a valuable role for the church in society. One gets the impression that Lanier wants people to quit both their social media accounts and their church.

One could do without some of the colorful language used in parts of the book, but the language does not diminish Lanier's arguments. The book convincingly warns its readers of the destructive effects of social media on individuals and society. It is timely for both thought and action.

*Reviewed by Michael Janzen, Associate Professor, Department of Computing Science, The King's University, Edmonton, AB T6B 2H3.*

**EVERYBODY LIES: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are** by Seth Stephens-Davidowitz. New York: Dey Street Books, 2018. 352 pages. Paperback; \$16.99. ISBN: 9780062390868.

*Everybody Lies*, as the subtitle suggests, is Seth Stephens-Davidowitz's book about "Big Data" and what it tells us about ourselves. He is quite explicit that he is inspired by *Freakonomics* and hopes to apply its irreverent but quantitative approach to new kinds of data that have been enabled by the internet.

Stephens-Davidowitz is an economist; however, his choice of topics strikes me as being rather sociological. More important to his work are his data sources. Key sources include search strings in Google, search strings used at the website Pornhub, Facebook posts, tweets, word frequency counts from Google Ngrams, and more. Stephens-Davidowitz makes a good case that this sort of data, only recently available to the public, has been underused by researchers. However, while some of this avoidance stems from a lack of creativity, there are many statistical issues (such as self-selection, nonrandom groups) that make it hard to meet academic standards with these sources.

Much of the book looks at the search strings used to find things on the internet, and these provide a good example of the challenges of these new data types. The companies the author works with to obtain this data provide him the anonymized strings with some counts on how often it occurred and possibly certain background data such as geographic area or gender. However, the statistics about the search strings are less helpful if we cannot relate them back to a population that we understand. There is no guarantee that internet users are “representative” of the population.

Unlike survey questions, in which everyone gets the same questions and answers, everyone gets to type in their own search string. This leaves the researcher on her own to decide when two different strings are “equivalent” and thus should have their counts combined. Stephens-Davidowitz makes a good case that the same terms in a different order can make a difference in meaning. For example, looking at searches with the names of candidates for political office, say A and B, a search string with A first and then B shows a preference for A, whereas a search string with B first and then A shows a preference for B, possibly even an unconscious preference. Even if one accepts this case, how is one to generalize it? Does the order matter for two different competing products? Does the order of my grocery list matter? It might at first appear that we must accept any difference as significant, but that probably gives us too many different categories from which to draw conclusions. However, trying to combine multiple terms into one category gives us a problem with researcher bias. If nothing else, the researcher has to assume a particular understanding of what the user really means, even when it is expressed differently.

We have counts for the search strings, but this does not mean that the count represents unique users. For example, if one has a simple situation in which the search string is A or B, it is possible that five users do search A and one user does search B five times. Based on the count, A and B are equally likely, but

if I care about the underlying population of users, a particular user is five times more likely to search for A than for B. The number of times someone makes the same search would seem to be associated with a particular user, not random. For instance, the address I am most likely to search for in google maps is my home address, as a starting or ending point.

Many of the topics that interest Stephens-Davidowitz are those in which people tend to avoid the truth; hence the book’s title *Everybody Lies*, or at least they lie on surveys. The topics in the book include sex (quite a bit), race, cheating on taxes, and more. The difficulty with these topics is well enough known to have its own technical name: social desirability bias. This is a bias in which people answer questions in a manner that will be viewed favorably by others, a form of hypocrisy. As La Rochefoucauld said: “Hypocrisy is the homage which vice renders to virtue.” While inconvenient for social scientists, this is an inevitable consequence of having a conscience—even if badly damaged—in a fallen world. People retain a sense of what is and what should be.

In an effort to work around the inaccuracies caused by this bias, the author looks for sources in which someone voluntarily discloses information, which, in his work, is often a search string. He may have found a way around the problem, but such remains unclear. When the user enters a search string, it is voluntary, and the string is one of their own choosing. It is unprompted by a survey, and it is anonymous. This appears to avoid social desirability bias; even so, there is no reason to think that we have avoided a self-selection problem. The very approach we use to avoid social desirability bias, that of a user voluntarily picking a search string, means that the user is self-selecting. The social sciences have long been concerned about self-selection and have been dubious of studies that fail to account for it.

*Everybody Lies* succeeds, in the spirit of *Freakonomics*, in telling some good stories that tie back to quantitative thinking. Stephens-Davidowitz shows creativity in finding information from new data sources. However, this often takes us into areas where we do not understand the data well. A common problem with his work is a desire to delve into areas involving social desirability bias, areas that people are reluctant to talk about. In trying to handle this, he almost certainly strays into the problems of self-selection, which makes his samples unrepresentative and, in turn, makes it difficult to draw valid conclusions. While *Everybody Lies* opens up vistas of new possibilities, its explanatory reliability is questionable.

*Reviewed by John Hunt, Professor of Computer Science, Covenant College, Lookout Mountain, GA 30750.* ◊