Vullioud, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior, 38*(1), 9–17.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686.

A NEW PARADIGM FOR BEHAVIORAL SCIENCE

Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are

By Seth Stephens-Davidowitz. New York, NY: HarperCollins, 2017. 352 pp. Hardcover, $27.99.

In a previous review, I described Chris Chambers's (2017) condemnation of many current practices in experimental psychology. These practices include seeking only results that support the investigators' biases, or at least changing their hypotheses to bring them in line with the results. This strategy is easily achieved by making questionable changes in the experimental procedure and misusing inferential statistics. These habits depict a fairly pessimistic expectation of how experimental inquiry might advance our accumulation of an understanding of behavior. Even with his recommendations for reform (which are being implemented, e.g., the Peer Reviewers' Openness Initiative), one might become disillusioned with experimental psychology. This paradigm held the promise of formalizing general laws of behavior based on impeccably controlled experiments and analyses. And, certainly, we have documented a few principles that seem to hold up fairly well in repeated experiments. These include Weber's law, the power law of practice, Shepard's law of generalization, and the Rescorla–Wagner law of learning. Although these laws are discernible in controlled investigations, it is somewhat difficult to measure their influence on behavior in the real world. This limitation should not dampen our enthusiasm for seeking general laws, but perhaps another complementary paradigm can make significant inroads into the understanding of behavior.

In fact, I was just reviewing the galley proofs for my book review of *The Seven Sins of Psychology* when Seth Stephens-Davidowitz's book, with a foreword by Steven Pinker, arrived. The author trained as an economist, worked for Google, and is currently a journalist for the *New York Times,* and he promotes the value of big data and data that are novel to behavioral science. A number of lessons are pounded home in the book, which germinated from the author's dissertation but also includes many other studies relevant to his main thesis that typical surveys do not truly capture our beliefs, our motivations, and how we behave. The guiding premise, well known to psychologists, is that we cannot believe what people say. Prototypical surveys are not going to cut the mustard with respect to predicting our behavior or even what motivates our behavior. Of course, psychologists have been demonstrating how misleading our impressions can be since the origin of our psychological science. Even if people do not consciously lie when asked, we know that we cannot trust what they say.

There are many different types of results that reveal the limitations of introspection. In one classic study, participants were asked to memorize word pairs, such as *ocean–moon* (Nisbett & Wilson, 1977). After this task, they were asked to name a detergent. Studying *ocean–moon* increased the likelihood of giving Tide as an answer, but when asked participants almost never mentioned the words in the memorization test as an influencing factor. In another task, participants chose an article of clothing of the best quality from several articles arranged in a row (Nisbett & Wilson, 1977). Participants revealed a strong position effect in that there was a strong bias to pick the nightgown on the right. Participants were four times as likely to choose the one on the right regardless of the actual nightgown in that position. When the participants were asked why that particular one was preferred, they never mentioned position as an influencing factor in their decision, and virtually all participants adamantly denied the experimenter's proposition that position had an influence. When I first studied probabilistic decision making in graduate school, the subjects would report that they used highly complex rules to decide their choices. But the stimulus and response on the previous trial was sufficient to predict their response to a given stimulus (Massaro, 1969). As documented in many different types of experiments, the introspective method falls short of providing an understanding of behavior in these domains. Introspective and even perceptual reports require sophisticated analyses, as witnessed by the contribution of signal detection theory and innovative tasks in behavior measurement.

Stephens-Davidowitz criticizes survey data for another reason. We tend to put a positive face on ourselves, our family, and our friends. This Pollyanna principle is of course conducive to positive behavior in contemporary society. But under this positive surface lurks a lot of negativity. This negativity is putatively revealed in Google search behavior. A Facebook post celebrates 7 years of a happy marriage while the husband asks Google, "Is my wife is cheating on me?" and his wife does a Google search of "Why won't my husband have sex with me?" For the author, digital truth comes from searches, views, and clicks or swipes, whereas lies are found in social media posts, social media "likes," and dating profiles.

Big Data

The new paradigm is big data, without depending on either introspective reports or survey results. Who would have believed that two engineering students from Stanford would create an enterprise that would produce a valuable source of big data in addition to facilitating Internet searches? It is difficult to imagine that Google now processes more than 40,000 search queries every second on average, producing about 144 million searches per hour, 24 hours a day, 7 days a week. Google has "kindly" made some of this information available to data scientists and anyone else with interest in this search behavior. The database is Google Trends (https://trends.google.com/trends/), which is an unbiased sample of Google search terms. Institutional review board approval is not necessary because the results are completely anonymous, but it includes the geographic locations of the participants. At this stage, the program does not make available age, sex, and other personal information.

The database for Google Trends comes equipped with an explorer tool to examine a search term or particular topic at specific times since 2004. The results are in the form of the proportion of all searches for that term at a given time. The results are normalized so that the metric gives a true ratio of the number of searches for that term divided by the total number of searches. The maximum number of searches for a given time and location is assigned the value of 100, and all other values are normalized accordingly.

Figure 1 gives search results since 2004 for "confirmation bias" on Google Trends. This figure shows fairly negligible interest until around 2008 and then a roughly linear increase until today. Given that we can expect searches to be carried out predominantly by nonprofessionals, it might be encouraging that a bias with empirical and theoretical foundation is becoming better known to laypersons. It would be nice to know how often people search for "confirmation bias" relative to searching for "pumpkin pie" or some other set of words that display interest. Apparently, Google does not like to disclose this type of information. Because we cannot get absolute numbers, we do not know how often people are really searching for "confirmation bias." The fairly large swings in search rate for "confirmation bias" across the different time periods might mean that the search term is fairly rare. In this case, it might be an instance in which big data in fact gets reduced to fairly small data.

Why should the Google Trends database be of particular interest? According to the author, the reason is that people are at their most honest when they are searching unattended and anonymously. They no longer have to put on a good face to the world, as they do on Facebook, for example. In the author's words, "I have come to believe that the new data increasingly available in our digital age will radically expand our understanding of humankind. . . . And new, digital data now shows us there is more to human society than we think we see" (p. 16).

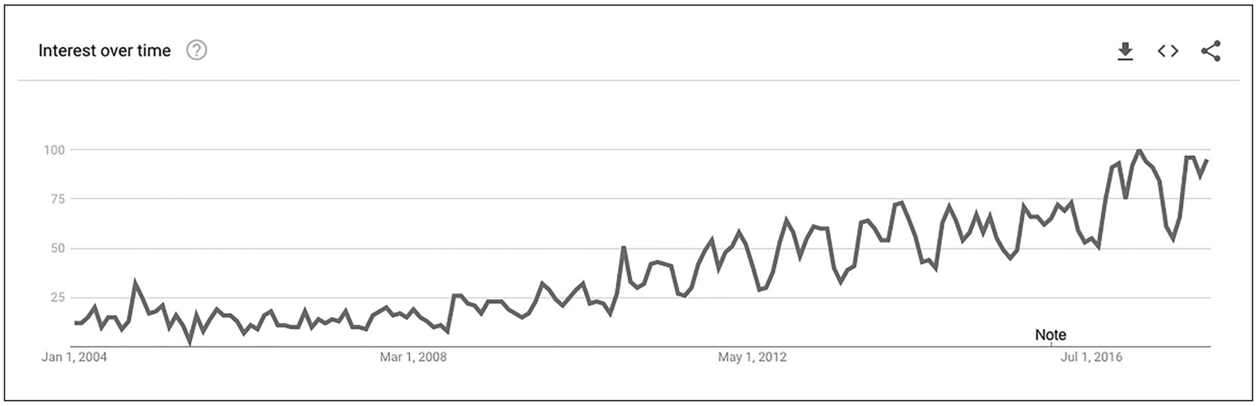

FIGURE 1. Google Trends result for the search term (or search topic) confirmation bias

The database is also not constrained by our incapability to introspect about the causes of our behavior or by the ability to access unconscious influences on our behavior (Bargh, 2017). We can expect that big data will have even more of an impact as computational science advances in developing and applying algorithms to analyze huge caches of data. The Flatiron Institute (https://www.simonsfoundation.org/flatiron/) is a recent privately funded enterprise aimed at uncovering gold in the digital world across many different domains of inquiry.

Findings from Big Data

The subtitle of the book is "What the Internet Can Tell Us About Who We Really Are." Big data has informed the following questions.

A good approximation is that about 5% of the male population is gay. It is difficult to determine an analogous estimate for women.

We can be comforted by the convincing findings that advertising works, so perhaps we can be more accepting of being bombarded with ads. The effectiveness of promoting products ranging from Superbowl ads to Yelp reviews is substantial. As an example, a restaurant having one fewer star on Yelp reviews can see revenues decrease by 5%–9%.

We learned why American Pharoah, a Triple Crown winner, was a great racehorse because the analyst understood that there is no data like more data. Measurements of his internal organs revealed that they were at the top of the charts in size. His left ventricle was at the 99.61 percentile. The lesson to be learned is that the size of a horse's internal organs is important and that no measure should go untested. This measure in combination with many others evidently is successful in predicting winning race horses.

To answer whether the media are biased, we can measure the language they use, specifically what words they use. A big data resource is Google Ngrams. Using Google's ambitious project of converting all analog books to digital form, it is possible to determine precisely what words were used in books published in various years. I learned that the word *sausage* began occurring around 1800, increasing in usage until 1900 and leveling off until now. The word *pizza,* on the hand, did not occur until about 1950 and increased dramatically until the present day. As most convincingly demonstrated by George Lakoff (1989), Democrats and Republicans use different language. Guess which legislators use "death tax" and which use "estate tax"?

What about the media's use of language? The author details a comprehensive study by Gentzkow and Shapiro (2011), who looked at politically charged phrases in hundreds of newspapers. They found that newspapers service their readers: Geographic areas inhabited primarily by conservatives read local conservative-slanted news, and analogously for liberals. Overall, newspapers are slightly liberal leaning, yes, but the authors reason that, overall, readers of newspapers bend in the liberal direction.

Based on research by experimental psychologists, we know that Freudian slips of the tongue are not Freudian (Baars, 1992). Stephens-Davidowitz supplements these conclusions by analyzing 40,000 typing errors and demonstrating that there is nothing Freudian hidden in these errors. Furthermore, his mining of a database of reports of dreams comes up empty, so, for example, the fruits we dream of are not biased to be Freudian-shaped.

Big data provides many avenues of inquiry. It can shed light on who cheats on their taxes. Not dishonest people but rather those who are in contact with tax professionals or other cheaters are most likely to cheat. Social influence rather than personality appears to be the culprit.

Contrary to popular belief, success in life cannot be simply explained by where you went to high school or college. Yes, overall, successful people attend good high schools and prestigious colleges. But correlation is not causation. The author reviews two independent creative analyses of big data. In one, students just made or missed the cutoff for attending a hugely prestigious high school. In another, having been accepted to both a prestigious and a less prestigious college, some chose one or the other, for example, choosing the "lesser" college for personal reasons. Greater success later in life for these cohorts did not reveal those who just made the cutoff for high school, or likewise for students who chose the more prestigious college.

Nowhere does big data play a bigger role than in the stock market. But forget about making a killing in the stock market now that you know about big data. Rest assured that a stock's current price reflects lots of big data. You are unlikely to find the secret predictor because so much inquiry has already occurred and because there is unlikely to be just one. Perhaps nothing illustrates a general principle of multiple influences on most phenomena more than the stock market.

Thinking about where to raise your kids can be informed by big data. How about a little peaceful farm or at least a suburban home surrounded by open space? The author downloaded Wikipedia and looked at 150,000 Americans with Wikipedia pages, assuming these entries reflected some form of success. Analyzing their backgrounds led to the author's conclusion, "These are just correlations, but they do suggest that growing up near big ideas is better than growing up with a big backyard" (p. 184). A control group of non-Wikipedians might have been more reassuring at least in terms of allowing some quantitative assessment of the importance of the geographic area of a child's upbringing. Or perhaps some simple ranking of the Wikipedians could be determined (such as how often their pages are accessed) along with a measure of urban-ness in order to derive some correlation between these variables. Before you move your family into the city, however, know that other variables such as living embedded with immigrants engenders prominence.

Big data with transcripts from speed dating encounters can also inform what a man should talk to a woman about if he wants a repeat encounter. Do not ask mundane questions about hobbies, films, or commuting. As Leonard Cohen advised his son, be a good listener and laugh at her jokes, stick with topics she offers, and feign interest in her interests. Apparently, women and not men tend to elongate the vowel *o* in the word *so* when they are interested in the fella. Yet the best strategy has not been determined. As I just heard on *Wait! Wait! Don't Tell Me,* just mentioning *guacamole* doubles the interest one gets in speed dating.

Perhaps the most dramatic effects of big data concern racial prejudice. Social psychologists have long studied the issue with explicit measures, but they were not predictive of prejudicial behavior. The topic received renewed attention with the invention of the Implicit Association Test (IAT). This test is aimed at revealing implicit unconscious biases by simply having participants choose one of two responses to various words and faces. For example, for the race IAT, the two categories might be "good and white people" or "bad and black people." Another condition would have the opposite pairing: "bad and white people" or "good and black people." A racial prejudice would putatively occur when a white person takes more time to categorize a white face as "bad and white people" than as "good and white people." An analogous prejudice would occur when a black person takes more time to categorize a white face as "good and white people" than as "bad and white people." This result has generated a huge amount of controversy in terms of reliability of the findings, their theoretical interpretation, and the consequences for society (see Mitchell & Tetlock, 2017, for a comprehensive and critical devastating review). Search data documented by Stephens-Davidowitz put a new light on measuring racial prejudice. Google search data revealed that racist terms and phrases were rampant, and they distributed themselves systematically nationwide. In the 2012 election, parts of the country with a large number of racist searches gave fewer votes for Obama than for John Kerry in the previous presidential election. Areas of our country that gave the most support for Trump were those that made the most Google searches for the n-word. We know that correlation does not prove causation, but it is a sobering finding nonetheless.

Big data can also inform what makes a story go viral. Two scientists used sentiment analysis to measure *New York Times* news stories. Sentiment analysis measures the average mood of the story by the overall positivity or negativity of the words in the story. Contrary to popular opinion, the numbers of stories most likely to be e-mailed indicated that positive stories outscored negative stories. Have a happy message if you want people to read your work.

Confirmation bias reinforces our beliefs, and certainly how much time we spend on our popular news sources reflects that bias. However, our time on the Internet tends to cast a wide net, so our reading might be more eclectic than we imagine. Gentzkow and Shapiro (2011) found that being online rather than offline increases the likelihood that you will encounter someone with opposing views.

Perhaps best of all, big data offers to build you a respectable wine cellar. The author piqued my interest with Orley Ashenfelter's First Law of Viticulture for wines from the Bordeaux region. Having 30 years of weather available, he was able to find high-quality wines by using winter rainfall, average growing season temperature, and harvest rainfall as predictor variables. You will have to look at the book or find the original source to find the coefficients of these three variables. Of course, soil and grape varieties are also important variables, but the assumption must have been that these did not vary much for wines that were analyzed.

How does one carry out experimental interventions in the domain of big data? You do not unless you are Google or some comparable entity. If you are, you do what is called A/B testing. By assigning two different search page styles to different users, for example, you measure their relative ease of use, as measured by clicking or swiping behavior. Of course, there is a tremendous advantage of A/B testing when so many users participate in the game. An answer to a question can be quickly obtained.

Limitations of Big Data

Envision with me the following scenario. You are a fresh PhD in behavioral science, and your new job offer promises you a huge grant to obtain an organized and comprehensive database of trillions of behavioral events by millions of different individuals. You will not need any additional grant support because this resource is relevant to almost any question you envision, is easily searched and represented, and provides a detailed and truly representative answer to your question. Your employer adds that, by the way, the database has already been available for a few years, and you can begin your inquiry immediately. How can you say no? So is big data, particularly in the guise of Google Trends, a revolutionary breakthrough in behavioral science? A few observations are relevant.

Stephens-Davidowitz descends back to Earth with his discussion of the limitations of big data and its potential pitfalls. Not unlike the challenge faced by traditional experimental psychologists is the plethora of variables to test. He calls this problem with big data the curse of dimensionality. The curse of dimensionality runs counter to the mantra that there is no data like more data. With excessively complex databases, there are just too many potential variables to test. With so many multiple tests, one or even several are likely to be significant. Enough monkeys at keyboards will eventually create an acceptable work of literature. This is analogous to the danger of multiple inferential statistical tests in traditional experiments. The author describes the misleading journey of putatively locating DNA causes of geniuses. It is not so simple even with big data.

The final word seems to be that multiple sources of influence or information are responsible for our behavior. There are sophisticated techniques for testing for the effectiveness of many variables in large databases, and investigators should pursue these in their inquiry. For example, the author describes how evaluating educational performance is best accounted for by a combination of test scores, student survey, and teacher observations. The three measures together gave the best results for evaluating successful education. We have arrived at similar conclusions about the multiple variables that influence a child's acquisition of spoken language. Parental input, imagery (concreteness), iconicity, and difficulty of articulation all have independent influences on vocabulary acquisition, and these influences change systematically across development for both receptive and productive language (Massaro & Perlman, 2017).

I believe that expanding our toolbox for inquiry can only improve our understanding of behavior. The author reviews a study by three economists that addressed how likely people who borrow money are to repay the loan. They looked at the written language potential borrowers used to describe why they needed a loan in addition to their credit ratings and income. They found specific phrases that correlated both positively and negatively with paying back the loan. It is instructive that the phrases *debt-free, minimum payment, lower interest rate, graduate,* and *after-tax* positively predicted paying back the loan, whereas *God, promise, will pay, thank you,* and *hospital* negatively predicted paying back the loan. Although the author agonizes that banks' use of this information might have ethical implications, I do not see any real problem with using this information relative to the other information such as defaults on previous loans. On the other hand, if language use predicted IQ, it might be a borderline case to use the language in a mandatory writing assignment for hiring or promotion practices. Finally, the author cautions that secluded Google searches cannot predict individual behavior, and it would be wasteful and unethical to intervene in person's life because of their Google searches.

The author's presentation puts a damper on the belief that we will be able to uncover general principles of our behavior in psychological inquiry. The mining of big data appears to reveal that the outcomes that are observed are highly situation specific and therefore would have applications only in those specific areas and not be easily generalized to other areas. It is perhaps relevant that a similar conclusion holds for artificial intelligence: The most convincing feats of artificial intelligence involve highly constrained feats, such as playing Go, driving a car, or optimizing an Internet search. Therefore, although big data might make it easier to uncover general behavioral laws, given that they are highly situation specific, we will at best have an encyclopedia of behavioral tendencies in specific situations rather than some general principles that are found in biology, chemistry, and physics, for example.

Finally, the subtitle of the book includes the statement "What the Internet Can Tell Us About Who We Really Are." Are secluded searches on the Internet any more of valid picture of who we are than when we wear our societal face among our friends or our vocational face among colleagues? As social psychologists would advise, the question for data scientists might be to ask what self we take into the voting booth, what self we present to family and friends, and what self we wear at work (Neisser, 2006). This question might be hard to answer because surveys, introspective reports, and even Google searches do not necessarily predict how we might behave differently in the many roles we play.

Conclusion

The author offers a creative solution to his concluding chapter by first confessing that there is no storybook ending about how data science improved his life, as it might do for all of humanity. He takes solace in the finding that only 7% of readers finished an electronic copy of Kahneman's (2010) book and only 3% finished Piketty's (2014).

Using his fandom for the New York Mets, the author reinforces a hypothesis of mine that many aspects of our personality, including our beliefs, are implanted in us at a very young age. His case for an early adoption of favorite sports teams can also be made for moral and political values. A supporting case study is an account of how seamlessly one can grow up racist, making it difficult to change this stance in adulthood (Pettigrew, 2018). If we combine these early beliefs with a tendency for confirmation bias, then we have an explanation for individuals who differ dramatically in their beliefs, which are not easily changed across the life span. As mentioned previously, social media on the Internet persuade us to be more open to reading posts and literature from different points of view. So perhaps there is a chance for a more open yet rational understanding as we carve our path ahead. The engaging and informative book by Seth Stephens-Davidowitz is a good road to continue our journey.

Dominic W. Massaro
Psychology Department
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
E-mail: Massaro@ucsc.edu

NOTE
Appreciation to Bill Rowe for his, as always, thoughtful ruminations on this review.

REFERENCES

Baars, B. J. (1992). *Experimental slips and human error*. New York, NY: Springer.

Bargh, J. (2017). *Before you know it: The unconscious reasons we do what we do*. New York, NY: Touchstone.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton, NJ: Princeton University Press.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics, 126*(4).

Kahneman, D. (2010). *Thinking fast and slow*. New York, NY: Farrar, Straus and Giroux;

Lakoff, G. (1989). *Don't think of an elephant: Know your values and frame the debate.* White River Junction, VT: Chelsea Green Publishing Co.

Massaro, D. W. (1969). A three state Markov model for discrimination learning. *Journal of Mathematical Psychology, 6,* 62–80.

Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers Communication*. Retrieved from https://doi.org/10.3389/fcomm.2017.00004

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a proxy for utility: The case of implicit prejudice. In S.O. Lilienfeld & I. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions*. Oxford, UK: Wiley.

Neisser, U. (Ed.). (2006). *The perceived self: Ecological and interpersonal sources of self knowledge* (Emory Symposia in Cognition). Cambridge, UK: Cambridge University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Pettigrew, T. F. (2017). *Growing up racist*. Book review of *The making of a racist: A southerner reflects on family, history and the slave trade*, by Charles B. Dew. *American Journal of Psychology, 130,* 521–523. doi:10.5406/amerjpsyc.130.4.0521

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: Harvard University Press.

HOW YOUR SMARTPHONE WAS ENGINEERED TO OUTSMART YOU

Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked

By Adam Alter. New York, NY: Penguin Press, 2017. 368 pp. Hardcover, $27.00.

How many times a day do you think you use your smartphone? 100? 1,000? According to one study, the *average* smartphone user swipes 2,617 times a day, and the top 10% of smartphone users clock in at an eye-popping 5,427 times per day. That's an average of once every 33 seconds (and once every 16 seconds for those in the top 10%).