Data Mining: The Future of Psychological Research… or a Runaway Train?

Morgan E. Gleason

State University of New York at New Paltz

A review of: *Dataclysm: Who We Are (When We Think No One's Looking).* By Christian Rudder, Crown, 2014. ISBN: 978-0385347372

In today's tech-savvy culture, we have become a society that is perpetually updating our Facebook or Twitter accounts, repeatedly searching on Google, managing our finances online, and texting and sending photos to family and friends – for starters. If being "plugged in" is the new norm, the result is two-fold: a more connected society, but also one that constantly, whether voluntarily or not, transmits personal data across the wires. Can this information be utilized conscientiously to better understanding humankind? In the thought-provoking *Dataclysm: Who We are (When We Think No One's Looking)*, Christian Rudder touches on these and other issues. Although the book only makes a few allusions to work from the field of evolutionary studies, examples of evolution at work in both the mating game and modern social interactions abound in this fascinating and data-oriented, yet down-to-earth, treatise on modern social behavior.

Christian Rudder, a Harvard graduate and one of the founders of the popular internet dating site OKCupid, attempts to address some of these new concerns in *Dataclysm*. At the start, Rudder explains how both his access to and ability to analyze data from millions of users puts him in the unique position to address some of these modern questions. He also manages to tell some attention-grabbing stories about human interaction itself. The book also inevitably raises the obvious questions about rights to privacy for users. Given the underlying empirical approach that drives Rudder's work on data, its important to note that Rudder's analysis is largely atheoretical. The evolutionary approach, which I myself am providing as a reviewer, has the capacity, I believe, to provide a vivid and coherent framework for many of the findings that Rudder describes.

AUTHOR NOTE: Please direct correspondence to Morgan E. Gleason, State University of New York at New Paltz, Department of Psychology, 369 Crescent Ave, Highland, New York, 12528.
E-mail: n00249271@hawkmail.newpaltz.edu

The author's primary goal is to rally support for the utilization of online data as a valuable research tool. The author stresses the usefulness of evidence garnered from the internet, maintaining that it provides researchers with a more representative and telling mosaic than most academic research that is conducted on a limited sample of college-aged students. Through the use of the Google Trends tool, Rudder manages to uncover secrets behind the ever-popular Google search bar. By analyzing data directly from Google's database, the author escapes social desirability bias and jumps straight into the search engines and, presumably, into the minds of the masses. Luckily, by sparing the reader the wonkiness of statistics jargon, the author presents his findings in a simple and evocative way, displaying results using graphs and images that a reader without a statistics background can easily comprehend. Along the way, he analyzes some thought-provoking behavioral patterns.

Rudder starts by discussing some curious findings from his OKCupid data, dealing with male and female dating preferences, common (and often comical) themes garnered from users of different cultures, and some interesting patterns specific to online behavior. He walks us through some rather non-intuitive findings in a clear, direct and entertaining way, making his findings accessible. The author makes a point, in each section of his book, to address the dangers of data mining and to acknowledge alternate perspectives in a judicious manner.

In Chapter 1, for example, he provides the reader with some intriguing data regarding the specifics of how online-daters converse on OKCupid. Here is where I witnessed the usefulness of evolutionary psychology as a lens with which to interpret his discoveries. The author discovered through researching the number of characters individuals typed in a message and the corresponding response rates that certain individuals were simply copying and pasting a formulated message to multiple potential "dates." This seemingly nonchalant approach was curious to the Rudder, who further wondered which method proved more effective. Is it better to take time and craft a specific message for each person you are interested in? Or to craft a blanket message, alter is slightly perhaps, and send it to more individuals in less time? The author was interested in exploring whether copying and pasting a template message was as effective as composing a more detailed and elaborate message. The author looked at the number of key-strokes that males utilized when messaging a potential female love interest. He discovered that 20 percent of his sample utilized this copy-and-paste method. As it turned out, the boilerplate messages were 75 percent as effective as writing something unique. Although initially somewhat surprising, this makes some sense when you consider effort-to-payoff ratio, a concept intrinsic to evolutionary mating strategies. By creating a template message, men have the ability to contact numerous women, and are expending less effort in the process, at little cost because they are still receiving replies from target females. By spreading their message to more potential dates, they improve the odds of obtaining a favorable reception. This particular behavior clearly illustrates how evolutionary

psychology pervades even modern online dating behavior. The evolutionary lens can help us better understand possible explanations for these types of findings – and, further, this kind of analysis can help lead to novel questions regarding the psychology of social interactions and relationships.

Another major emphasis in this fact-laden book is how the internet manages to rapidly create a mob mentality on social media sites like Twitter. This behavior can range from a collective anti-corporate screed, to downright scurrilous remarks aimed at fellow bloggers. It seems, for anyone in the limelight, that saying the wrong thing on Facebook or Twitter pretty much assures a public thrashing. The author points to several examples of this in the book. The concept of moral outrage and punishing behavior is well documented in the field of evolutionary psychology (Brown, 1991; Wilson, 1993). As Wilson (1993) points out, these behaviors play a crucial role in our social functioning – with the primary goal of preserving ethics within our social group. The internet provides a platform for regulation, by allowing users to visibly declare norm violations, in addition to a setting for carrying out public punishment. In evolutionary psychology, and on the web, punishment is the cost exacted on those who fail to adhere to the moral conventions of the in-group. Again, by using the evolutionary perspective to interpret his findings, more specifically the concept of moral outrage, can assist in better understanding these behaviors.

Having evolved, as a species, in large groups, it seems justifiable to say that competitive behaviors like status signaling and suggesting social influence would have been prominent. Even in the modern corporate world and across social media outlets, these behaviors are palpable. As Buunk and Ybema (1997) argued, flaunting high status and prestige serves to bolster self-esteem and indicates the health and competence of the in-group to which an individual belongs. Rudder also addresses this concept, albeit utilizing some contemporary jargon. In this modern age, he argues, it's frightening to think that going forward; we're all effectively "managing our brand" by posting on Twitter or Facebook, having to project a unique version of ourselves out to the internet in order to be efficacious.

Although it's true that data in the aggregate aren't so frightening, especially if there is no personally identifiable information in a data set, in fact, the author argues, it can lead us to a better understanding of ourselves. Rudder urges us to recognize that our seemingly private behavior – the things we search, our messages to friends, and our online dating behavior – are in fact public. The author reminds us that that is the very reason sites like Google and Facebook are free. Rudder doesn't offer a simple solution to this problem (I'm not sure there is one), but suggests that the storm is still brewing. He does leave us with the sentiment that garnering something positive out of the data is worthwhile, and once we get a firmer grasp on the flood of information we will figure out how exactly to wield it. This book is scary, silly, eye-opening, and completely relevant for anyone in the modern age.

It was illuminating to discover how Rudder's work, which is

primarily about big data and the modern world, could provide so many great examples of the kind of research that can be elucidated by an evolutionary perspective. Rudder's work proved bursting with evolution-based themes from signaling status to mating strategies. Unwittingly, *Dataclysm: Who We Are (When We Think No One Is Looking)* provides a great example of evolution-based scholarship, with a modern twist.

REFERENCES

Buunk, A. P. & Ybema, J. F. (1997). Social comparisons and occupational stress: the identification-contrast model. In: Buunk, B. P. & Gibbons, F. X. (eds*). Health, coping, and well-being: perspectives from social comparison theory*, pp. 359-88. Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, D. E. (1991). Human universals. New York: McGraw-Hill

Wilson, J. Q. (1993). The moral sense. New York: Free Press