An Emerging Trend: Is Big Data the End of Theory?*

MICHAEL W. MACY

Abstract

"Big data" from online communities opens up unprecedented opportunities for social scientists to track human behavior and social interaction in real time and on the individual level yet on a global scale. Will the empirical windfall make theory unnecessary by allowing scientists to search for whatever pattern is "out there," expected or otherwise? This essay argues that the deluge of relational data promises a new beginning for causal explanation, for two reasons. First, these new sources of data—from friends, peers, neighbors, colleagues, and coworkers—avoid the atomistic theoretical bias imposed by a half century of reliance on surveys administered to stratified random samples. Second, the growing ability to conduct online experiments with randomized trials makes it possible to test theories about the causal processes that underlie observed patterns.

For the past century, social science has suffered from a severe and unrelenting empirical drought. The reason is simple: social interactions are very hard to observe and even harder to study under controlled conditions. People prefer to interact in private, they change their behavior when they know they are being observed, the number of dyadic interactions increases exponentially with group size, and these observations need to be ongoing—a single snapshot is often insufficient. These constraints have limited studies of ongoing behavior and social interaction to small bounded groups such as clubs (Zachary, 1977) and villages (Entwisle, Faust, Rindfuss, & Kaneda, 2007). Population scale data are limited to surveys because it is much easier to ask an isolated individual about their friends than to observe the ongoing interactions that comprise friendship. Surveys provide detailed retrospective accounts of individual attributes but are largely incapable of measuring real-time ongoing behavior, social interactions, or network structures.

*This essay incorporates material previously published in Golder and Macy (2014), Macy (2015), and DellaPosta, Shi, and Macy (2015).

Emerging Trends in the Social and Behavioral Sciences.
Robert Scott and Marlis Buchmann (General Editors) with Stephen Kosslyn (Consulting Editor).
 ISBN 978-1-118-90077-2.

BIG THEORY IN THE ERA OF BIG DATA

The empirical drought is now over. "When it rains, it pours," and social science is now flooded with a tidal wave of time-stamped digital records that track ongoing human behavior and social interaction at the individual level yet on a global scale, as Scott Golder and I noted in a recent overview for the *Annual Review of Sociology*.

The Web sees everything and forgets nothing, each click and key press residing in a data warehouse waiting to be mined for insights into behavior, from spam detection to product recommendations to targeted advertising. Our mobile phones, tablets, and laptops report every webpage we visit and every link we click and can even report our precise location and movements. Our social interactions are mediated through e-mail, Skype, instant messaging, Facebook, and Twitter. Our photographs are identity-tagged, geo-tagged, and time-stamped, creating a who-when-and-where recording of everything we upload. Social media platforms such as Facebook and online labor markets such as Amazon Mechanical Turk make it possible to conduct controlled experiments using thousands of participants from all over the world.

Meanwhile, rapid advances in cluster computing make it feasible to store and analyze these data at petabyte scale. As a consequence, it is now possible for the first time to search randomly for empirical regularities instead of targeting the collection of data needed to confirm a previously hypothesized prediction.

This ability led *Wired*'s Chris Anderson (2008) to famously proclaim "the end of theory." The end of theory is also embraced as the end of explanation: When the world can be mapped with unprecedented temporal and spatial granularity yet at global scale, it is sufficient to answer "how" and leave "why" to the storytellers. Similarly, we do not need to know why a prediction holds so long as it has high precision and recall.

A recent special issue of the *Review of Behavioral Economics* took a very different tack. In the lead article, Herb Gintis (an economist) and Dirk Helbing (a sociophysicist) pointed to the need for an "analytical framework for modeling the structure and dynamics of modern societies" (2015). In a comment on their proposal, I characterized the framework they propose as "big theory" (Macy, 2015). This framework uses recent advances in complexity science to go far beyond the equilibrium models that have been the mainstay of neoclassical economics for the past century. Gintis and Helbing identify two important theoretical breakthroughs: from equilibrium to disequilibrium and from instrumental rationality to normative and expressive behavior. These twin extensions offer the promise of intradisciplinary theoretical integration and greater opportunities for cross-disciplinary collaboration between social science and statistical physics. The intellectual forebear is Parsons and Shils (1951) general theory of social action, a midcentury attempt at theoretical integration that failed, Gintis and Helbing contend, because Parsons was working prior to the development of the powerful new analytical tools in complexity science, including stochastic and Bayesian updating, computational modeling, and evolutionary game theory. A half century later, all the pieces for "big theory" are now in place, thanks to theoretical breakthroughs that came too late for Parsons.

Aiming the Telescope

So who is right, Anderson, who sees the empirical and computational advances supplanting the need for theory, or Gintis and Helbing, who see advances in big theory as a new foundation for empirical research? Similar to Gintis and Helbing, physicist-turned sociologist Duncan Watts (2011) also sees exciting opportunities for social science, but for a very different reason. For Watts, the immediate stumbling block is the need for better data, not general theory, and the breakthroughs are empirical, not theoretical. Watts (2011) concludes, "[T]hree hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin."

Gintis and Helbing might quickly remind us that it is not enough to build the telescope. We also need to know where to point it, and for that we need the core analytical toolkit that they propose. The argument goes back to Parsons: "We never investigate all the facts which could be known about the phenomena in question, but only those which we think are important. This involves a selection among the possible facts," Parsons argues, based on "their relevance to the logical structure of a theoretical scheme" (1938, p. 15).

Or is it the other way around? Could Newton have discovered the laws of motion, Watts asks, without the prior invention of the telescope and the data collection it made possible? Parsons failed, Watts contends, because he did not have sufficient data to detect the behavioral regularities required for a general theory. A telling example: The Arecibo Observatory, devoted to the search for signs of extraterrestrial intelligence. The search is not limited to locations suggested by a theory of where to look. In addition, we now have the ability to systematically scan the heavens looking for evidence.

Yet Watts stops well short of embracing the end of theory. Instead, he recommends what Parson's student Robert Merton (1949) called *middle range theorizing*. Middle range theorizing can be situated between two extremes. On one side is a unifying theory that can be universally applied, such as Parson's general theory of social action or the analytical Marxism I learned from

Gintis when I was a graduate student. At the other extreme is the atheoretical practice of data mining and pattern detection advocated by Anderson. In between these extremes, one finds middle range theories that treat a particular empirical puzzle as a vehicle for addressing an underlying theoretical question.

The Big Difference

I agree with Watts but would take the argument one step farther: *The era of big data is not the end of theory, it is the beginning.* It is the beginning in the familiar epistemological sense that Kepler's observations enabled Newton's generalizations. Breakthroughs in measurement technologies often precede advances in theory.

Of course, this correlation is little more than a mantra unless we can spell out the enabling mechanism. Big data is the beginning of theory in a very specific way that goes largely unrecognized. What is important about big data is not just that it is big, but that it is different. The key difference is that the data are relational and behavioral, in striking contrast with the atomistic and retrospective self-reports that have been the mainstay of social science following the invention of stratified random sampling.

It is widely recognized that samples differ from the underlying population because of underrepresented, hard-to-reach subpopulations, and there have been enormous advances in the ability to overcome that problem both in the administration of surveys and in the statistical weighting of the results. There is another way that samples differ with the underlying population; however, that is not widely recognized. Respondents have no family members, neighbors, friends, colleagues, coworkers, or classmates in the sample. There is no one else in the sample who influences the respondent and no one whom they influence.

This is deliberate. Sampling frames are intentionally designed to obtain independent observations for two important reasons. Independent observations are used to estimate standard errors and to obtain samples that are representative of the underlying population distribution.

The problem is that the independent observations obtained using opinion surveys administered to a random sample imposes an atomistic theoretical bias. Where do we look for our explanations of the opinions we observe in the data? The search for explanations is constrained by the available data. We have no data about the opinions of the respondent's friends, coworkers, family, or neighbors. However, in most opinion surveys, we have very complete data about the individual attributes of the independent respondents. We know each respondent's socioeconomic background, ethnicity, gender, age, education, occupation, and income. We can then show how the distribution of opinion across a set of individuals is associated with the distribution of their demographic profiles. Why is a given individual more liberal on some hot button social issue? Because of that individual's age, gender, ethnicity, education, occupation, and income.

Not only do we have the data for atomistic explanations, the data are almost certain to cooperate in the empirical conspiracy. In a recent study, DellaPosta *et al*. (2015) found that 58% of the lifestyle measures in the General Social Survey (GSS) cumulative file were correlated with political ideology, and a follow-up analysis revealed that 94% of lifestyle measures were correlated with one or more demographic measures at the 0.001 level. Pick an opinion, attitude, or belief at random and you are very nearly guaranteed to find a highly significant demographic correlate.

In addition, demographic measures allow us to make convincing causal arguments using observational data, without the need for costly, time-consuming, and generally impractical reliance on experiments with randomized trials. Correlations between most other measures available to social scientists are susceptible to what statisticians call *spurious causation* (e.g., property damage tends to increase the more fire trucks that show up to put out the fire). In contrast, most demographic measures do not have "causal priors" (e.g., there is nothing prior to gender or ethnicity that causes one to be male or white and also affects opinions and beliefs).

Moreover, these demographic explanations are reasonable and ideologically convenient. They are reasonable because the opinions that we hold—our beliefs, ideas, and attitudes—reflect our formative personal experiences and our material self-interests. The explanations also resonate with the enlightenment commitment to an individualistic ideology that people "think for themselves," as if they lived on a deserted island with no influence from anyone else. In short, the atomistic theory is logically sound (given that demographics are causally prior), empirically plausible, resonates with the Enlightenment belief in individual autonomy, and enjoys massive empirical support in survey-based opinion research.

There is, however, a potentially devastating problem, one that goes largely unnoticed: Atomistic explanations—including those anchored in demographic causal priors—may be entirely or largely spurious owing to network autocorrelation. Methodology textbooks invariably warn us about serial autocorrelation—an observation at time t is not independent of that same observation at time $t-1$. For example, fast food consumption and longevity both increased in the postwar era, but that does not mean we should eat Big Macs for breakfast, lunch, and dinner. Network autocorrelation is the relational equivalent. Just as an individual is similar to its earlier self, so too an individual is also likely to be similar to its network neighbors, and this can lead to spurious correlation if the autocorrelation is not accounted for in the causal model.

In the same study that showed widespread correlations among lifestyles, politics, and demographics, DellaPosta *et al*. (2015) used computer simulation to demonstrate this formally. They created an artificial world based on two widely observed principles: homophily and social influence. *Homophily* refers to the tendency to interact with those who are similar and to avoid those who are different. Influence is the tendency to become similar to those with whom one interacts. The two self-reinforcing processes generate network autocorrelation. They assigned the agents a static attribute corresponding to a salient demographic trait and dynamic attributes corresponding to opinions that could be influenced by demographics and by the opinions of others with whom an agent interacted. From a random start, agents became sorted into densely connected like-minded clusters that also differed demographically. If we were to administer a survey to a random sample, we would find widespread correlations among opinions and between opinions and demographics, just like we find in the GSS. We might reasonably conclude that opinions reflected the formative experiences and material interests associated with demographic differences. However, if we then control for the opinions of network neighbors—which we can do in this artificial world but not in real-life random samples—demographic differences in opinion diminish by an order of magnitude.

The simulation results point to the troubling possibility that thousands of widely published statistically significant demographic correlations with opinion may be largely spurious owing to the unmeasured effects of network autocorrelation. The effects are unmeasured because the data on network neighbors are not available in the GSS or most other opinion surveys administered in the past century. With no ability to rule out network autocorrelation as a source of spurious causation, we can simply pretend that the observations in the underlying population must also be independent, just as they are independent in the sample that was carefully constructed to be representative.

The circle is easy to close. Statistical inference depends on random samples composed of independent observations, atomistic explanations have elective affinity with underlying populations composed of independent individuals, and representative samples give us confidence that the demographic correlations that are all but guaranteed to be discovered in a sample accurately reflect the processes operating in the underlying population. In short, the atomistic explanation is accepted because it is intuitively plausible, ideologically resonant, the available survey data strongly support it, and the relational data are not available that might refute it.

At least not until now. Relational data is pouring in from everywhere around us. Wikis make it possible to observe in minute detail how globally dispersed work teams coauthor documents. Music download sites such as Last.FM let us track the birth of stars in social networks. Every day, the "blogosphere" generates billions of public records containing the opinions of hundreds of millions of friends and followers around the globe, on topics including politics, health, travel, gardening, real estate, books, fashion, beauty, education, sports, and law. Crowdfunding sites such as Kickstarter track the flow of investments in start-ups.

Relational data is not limited to interpersonal ties in social networks. We can also construct bipartite networks in which the links represent a cultural connection. For example, we can use copurchases on Amazon, colistening on Last.FM, cofollowing on Twitter, and codownloads on YouTube to find cultural links among books, bands, and buffoons.

In short, data from these and other online communities are now being used for a growing number of studies that confirm relational theories of social influence and homophily, with the results published in the leading scientific journals. Access to big data is not ending theory; it is expanding it, from atomistic to relational.

LIMITATIONS OF BIG DATA

Just as the electronic microscope and space telescope allow us to see molecules and distant stars, massive amounts of online data make it possible to detect effects that are much smaller than what could be previously observed. For example, Scott Golder and I used hundreds of millions of tweets to detect diurnal patterns of positive and negative affect that can be detected by the choice of words (Golder & Macy, 2011). The effects of emotion on word choice produce strikingly stable patterns, but the effect sizes are tiny—not because the emotions are small but because emotions have only very small effects on the word choices that are used as lexical indicators of changes in affect over time.

The power to see small effects also has a downside. Big data make it harder to find the needle in the haystack not because of all the hay but because of all the needles. For example, a modest correlation of $r = 0.17$ is not statistically significant ($p < 0.09$) with $n = 100$, yet a vanishingly small correlation of $r = 0.0014$ is highly significant ($p < 0.00001$) with $n = 10\,\text{M}$. This time the correlation is vanishingly small ($r = 0.0014$), yet it is highly significant, with $p < 0.00001$. More generally, with big data we can find needles that are incredibly small, but the question is, do we care? And to answer that question, we can no longer use statistical significance. The criterion we have to use is theoretical significance, and for that we need theory.

We also need theory for causal inference, as noted by Parsons (1938, p. 15). Researchers, Parsons observes, are rarely "content simply to state bald, discrete facts. They go beyond this to maintain the existence of relations of interdependence, causal relations, [and] an imputation of causal relationship cannot be proved without reference to generalized theoretical categories." Moreover, even when the researcher is content to state a descriptive pattern, there is nonetheless a theory implied by the choice of what to include and how to measure it. Thus, the alternative "is not as between theorizing and not theorizing, but as between theorizing explicitly with a clear consciousness of what he is doing with the greater opportunity that gives of avoiding the many subtle pitfalls of fallacy, and following the policy of the ostrich, pretending not to theorize and thus leaving one's theory implicit and uncriticized, thus almost certainly full of errors" (1938, p. 15).

While I agree with Parsons that causal inference requires a theory, even if it is only implicit, it does not follow that theory combined with data is sufficient for causal explanation. No matter the size, big data remain observational in nearly all cases, and observational data have inherent limitations for causal inference. The abundance of big data is enabling the era of measurement, not the era of explanation. We need to recognize and appreciate what these data can tell us and avoid the temptation to claim more than what is warranted. Computational social science is a young science, and so we should expect the focus to be on measurement, on detecting the patterns that we could not see before. Even if we cannot yet explain what we can now detect, it is important to be able to identify the patterns that need explanation.

These data have an additional limitation. While survey data are strong on demographics and weak on relational measures, data from online networks are strong on relational measures but weak on demographics. Researchers are making rapid progress in building face-recognition tools to impute age, gender, and ethnicity, but much work remains to fill in other demographic dimensions, especially education, occupation, and income.

EXAMPLES

When Nathan Eagle, Rob Claxton, and I used a nearly complete 1-month record of telephone calls to map the social network structure of the United Kingdom, this was the first time this had ever been done for an entire country: 360 million social connections among 65 million people (Eagle, Macy, & Claxton, 2010). Nevertheless, our study raised more questions than it answered. We used these data to test Granovetter's (1973) theory of the "strength of weak ties," which predicts economic advantages of connections to people outside one's immediate network neighborhood. That is one of the most influential ideas in all of social science, and our data made it possible for the first time to demonstrate this economic advantage at population scale. We found that communities composed of people with diverse social networks were economically advantaged. Yet we could not identify the underlying causal mechanism or even the causal direction. That remains to be done.

These same data revealed another puzzle as well. Granovetter assumed that bridge ties are weaker—less frequent interaction and lower emotional valence (e.g., trust)—compared to ties that are embedded within densely connected clusters. We found that the density of interaction between network neighbors increases with the proportion of mutual friends, exactly as Granovetter predicted. In addition, we found that the density dropped sharply as the range of the tie increased from two steps (neighbors with at least one mutual friend) to four steps (only the neighbors of the neighbors have at least one mutual friend). However, then something really strange happens. As the range (i.e., the network distance spanned by the tie) increases beyond four, the density of interaction starts to increase.

Recently, Patrick Park and I replicated this analysis with data from Twitter and found the same pattern. It is an important contribution to have detected this unexpected pattern, even though we cannot yet explain it. There are several possibilities: (i) Higher degree creates more shortcuts that reduce the range of each tie, and the more neighbors one has, the less time to interact with each one. (ii) Relationships with business and professional associates may entail frequent interaction even though they are outside one's circle of friends. (iii) Social and geographic mobility causes social ties to acquaintances to decay (e.g., the ties to all but the very closest of childhood and high school friends are likely to have disappeared by middle age). These are all reasonable hypotheses, but we are not yet able to find a definitive solution to the puzzle.

In a recent study, Bogdan State, Ingmar Weber, Patrick Park, and I used global social network data to see if online social media can bridge cultural boundaries in the offline world (State, Weber, Park, & Macy, 2015). The boundaries we tested were proposed by Samuel Huntington, who posited eight civilizations that have persisted over the course of recorded human history—the Western, Orthodox, Islamic, Asian, and so on. We wanted to see if social ties are more likely to form between people living in different countries when the two countries share the same cultural heritage. To find out, we used Yahoo's anonymized global e-mail logs, with the message content removed, and taking into account differences among countries in Internet access and in Yahoo's share of the e-mail market. The results provided strong evidence that cultural boundaries persist in online social networks, despite the "world wideness" of the Web.

In a follow-up study with Wei Dong, Kate Ehrlich, and Michael Muller, we used data from one of the largest multinational corporations to see if common organizational membership is sufficient to bridge those same cultural boundaries (Dong, Ehrlich, Muller, & Macy, 2015). This corporation operates an internal Facebook-like social media platform to connect its 300,000 worldwide employees. We used the logs to replicate the e-mail analysis, except that this time everyone is in the same organization. Will that be sufficient to bridge the cultural boundaries that Huntington proposed? The answer is no. The results were nearly identical to what we found with e-mail logs. So then we drilled down a little deeper. What if everyone is working together on the same task? To find out, we mapped the global network structure for the 15,000 people on sales teams. This time the cultural boundaries completely disappeared. People on sales teams were just as likely to interact across cultural boundaries as within.

These studies of national and global social networks are examples of how big data are shifting the theoretical focus of social science from atomistic multivariate analyses of the relationships among attributes within individuals to relational analyses of the relationships between individuals. These data also allow us to go beyond the retrospective self-reports in survey responses to measure highly granular real-time behavior at a global scale.

COMING ATTRACTIONS

These examples of studies using online data represent breakthroughs in measurement, not explanation. They are based on observational data whose explanatory limitations do not diminish with the number of observations. Yes, we can now measure things that would have seemed impossible a decade ago, but as exciting as this may be, it is also a source of immense frustration. Social scientists want to do more than measure, we want to explain.

The gold standard for explanation is the use of randomized trials. For the past half century, social and behavioral scientists have conducted experiments involving a handful of college sophomores engaging face-to-face in a physical laboratory. Generalizing from a few dozen relatively advantaged 19-year-olds to the broader adult population raises troublesome questions about the external validity of the findings. In response, researchers have begun to use "virtual labs" in which a much larger and more diverse set of participants interact online. For example, Dustin Chertof, Elisa Bienenstock, and I developed a multiplayer game called *Empires of Fortune* to study intergroup rivalry. The game drew on proven formats such as "Civilization" and "Farmville" and confronted players with the dilemma whether to contribute to team efforts to compete with a rival group. Experimenters can manipulate game parameters to test how resource constraints, group size, network structure, and social identity affect conflict and cooperation.

Online labor markets, such as Amazon Mechanical Turk, are also being used as virtual labs. Milena Tsvetkova and I used a virtual lab to look for the causal mechanisms that explain why people respond to the kindness of strangers by "paying it forward" (Tsvetkova & Macy, 2014). We created chains of generosity to strangers by recruiting hundreds of "Turkers" on Amazon's "Mechanical Turk" online labor market. We found that receiving help made participants more likely to help others, but observing help had the opposite effect, as predicted by theories of collective action.

Other recent studies have gone far beyond the several hundred participants that are feasible using Amazon Mechanical Turk. Facebook has collaborated with social and behavioral scientists to run field experiments involving millions of users to see if friends influence voting behavior (Bond *et al.*, 2012, pp. 295–298) and if emotions are contagious (Kramer, Guillroy, & Hancock, 2014, pp. 8788–8790).

These studies have sparked intense controversy regarding the ethics of online field experiments. These concerns apply to observational studies as well. Online research poses serious legal and ethical challenges on two levels: how to protect individual privacy and how to reconcile property rights of the companies that own the data with requirements for open access with which to replicate results.

A careful exploration of these issues goes beyond the purview of this essay. The focus here is on concerns that massive online data may lead to a deluge of empirical findings with little or no theoretical motivation. As Scott Golder and I concluded in a recent review (2012), "The unprecedented opportunity to observe human behavior and social interaction in real time, at a microscopic level yet on a global scale, is attracting widespread interest among scientists with the requisite skills to mine these data but not always with the theoretical background needed to guide the inquiry. Studies that identify patterns of behavior or map social landscapes invite dismissal as 'atheoretical empiricism,' but this may be shortsighted. These pioneering studies should instead be taken as evidence not of the most that can be learned from online research but of the vast opportunities that lie ahead for a new science of social life."

REFERENCES

Anderson, C. (2008). *Wired. The end of theory: the data deluge makes the scientific method obsolete*. Retrieved from http://www.wired.com/2008/06/pb-theory/

Bond, R., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 295–298.

DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do liberals drink lattes? *American Journal of Sociology*, *120*, 1473–1511.

Dong, W., Ehrlich, K., Muller, M., & Macy, M. (2015). Embracing cultural diversity: Online social ties in distributed workgroups. CSCW '16 *Proceedings of the 19th ACM conference on computer-supported cooperative work and social computing* (pp. 274–287). ACM New York, NY, USA ©2016. ISBN: 978-1-4503-3592-8. doi:10.1145/2818048.2835198.

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*, 1029–1031.

Entwisle, B., Faust, K., Rindfuss, R., & Kaneda, T. (2007). Networks and contexts: Variation in the structure of social ties. *American Journal of Sociology.*, *112*, 1495.

Gintis, H., & Helbing, D. (2015). Homo socialis: An analytical core for sociological theory. *Review of Behavioral Economics*, *2*, 1–59.

Golder, S., & Macy, M. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*, 1878.

Golder, S., & Macy, M. (2014). Digital footprints: Opportunities and challenges for social research. *Annual Review of Sociology*, *40*, 129–152.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Kramer, A., Guillroy, J., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*, 8788–8790.

Macy, M. (2015). Big theory: A Trojan horse for economics? *Review of Behavioral Economics*, *2*, 161–166.

Merton, R. (1949). *Social theory and social structure*. New York, NY: Simon & Schuster.

Parsons, T. (1938). The role of theory in social research. *American Sociological Review*, *3*, 13–20.

Parsons, T., & Shils, E. (1951). *Toward a general theory of action*. Cambridge, MA: Harvard University Press.

State, B., Weber, I., Park, P., & Macy, M. (2015). The mesh of civilizations. *PLoS One*, *10*, e0122543.

Tsvetkova, M., & Macy, M. (2014). The social contagion of generosity. *PLoS One*, *9*, e87275.

Watts, D. (2011). *Everything is obvious: How common sense fails us*. New York, NY: Crown.

Zachary, W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, *33*, 452.

RECOMMENDED READING

DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do liberals drink lattes? *American Journal of Sociology.*, *120*, 1473–1511.

Golder, S., & Macy, M. (2014). Digital footprints: Opportunities and challenges for social research. *Annual Review of Sociology*, *40*, 129–152.

Macy, M. (2015). Big theory: A Trojan horse for economics? *Review of Behavioral Economics.*, *2*, 161–166.

Watts, D. (2011). *Everything is obvious: How common sense fails us*. New York, NY: Crown.

MICHAEL W. MACY SHORT BIOGRAPHY

Michael W. Macy received his B.A. and Ph. D. from Harvard, along with an MA from Stanford. He is currently Goldwin Smith Professor of Arts and Sciences and Director of the Social Dynamics Laboratory at Cornell, with a joint appointment in Sociology and Information Science. With support from the National Science Foundation, the Department of Defense, and Google, his research team has used computational models, online laboratory experiments, and digital traces of device-mediated interaction to explore familiar but enigmatic social patterns, such as circadian rhythms, the emergence and collapse of fads, the spread of self-destructive behaviors, cooperation in social dilemmas, the critical mass in collective action, the spread of high-threshold contagions on small-world networks, the polarization of opinion, segregation of neighborhoods, and assimilation of minority cultures. Recent research uses 509 million Twitter messages to track diurnal and seasonal mood changes in 54 countries, and telephone logs for 12B calls in the United Kingdom to measure the economic correlates of network structure. His research has been published in leading journals, including Science, PNAS, American Journal of Sociology, American Sociological Review, and Annual Review of Sociology.

RELATED ESSAYS

Agency as an Explanatory Key: Theoretical Issues *(Sociology)*, Richard Biernacki and Tad Skotnicki

To Flop Is Human: Inventing Better Scientific Approaches to Anticipating Failure *(Methods)*, Robert Boruch and Alan Ruby

Repeated Cross-Sections in Survey Data *(Methods)*, Henry E. Brady and Richard Johnston

Ambulatory Assessment: Methods for Studying Everyday Life *(Methods)*, Tamlin S. Conner and Matthias R. Mehl

Heterarchy *(Archaeology)*, Carole L. Crumley

Meta-Analysis *(Methods)*, Larry V. Hedges and Martyna Citkowicz

Ethnography in the Digital Age *(Methods)*, Alan Howard and Alexander Mawyer