

NGSA – Next Generation Social Analytics: Challenges and Payoffs

Special Session along with HUSO 2016, November 13 - 17, 2016 - Barcelona, Spain
<http://www.iaaria.org/conferences2016/HUSO16.html>

Dennis J. Folds
Georgia Institute of Technology
Atlanta, Georgia

Abstract— Social science stands on the brink of a revolution – or of failure. It needs powerful new tools, methods, and paradigms in order to succeed. These will include advances in computational capabilities, machine-based knowledge assimilation, quantitative analysis, and measurement. The future of social science could be exciting, as it studies the phenomena that are at the heart of the human experience. Failure to adapt, however, will lead to increasing irrelevance, reduced resources, and loss of status in the community of science. A special session on Next Generation Social Analytics is included in the Human Social Analytics (HUSO) 2016 conference, held in Barcelona, Spain. Two papers address the spectrum of challenges faced by social science and describe tools and methods that are key to success.

Keywords— Computational social science; social analytics.

I. A DIRE SITUATION

Social science is under intense scrutiny from politicians, funding agencies, leaders of scientific societies, and from within the constituent disciplines [1][2]. Well-publicized instances of fraud and misconduct may dominate popular headlines, but underlying problems with little replication, poor generalizability, poor methods, and biased interpretation of results are larger problems, and do not involve malfeasance.

Despite decades of studies of human behavior at the individual and group levels, comprehensive theories that adequately account for behavior in real world conditions remain illusive. Behavior is indeed complex, but at the root of social science is the conviction that behavior is lawful. Much of basic psychology (sensation and perception, for example) is well established. But studies of such constructs as beliefs, political action, and organized violence lack unifying theories that have any success in accounting for wide ranges of social phenomena.

Study after study is trumpeted in the popular media, as long as it fits the ideological preferences of the media gatekeepers, despite ongoing lack of replication and obvious failures in generalizability. Whether these shortcomings are largely the result of poor methods, poor interpretation, or simply the complexity of the phenomena studied, is not known.

Many other sciences, notably astronomy and biology, have come under intense scrutiny and criticism when results contradicted the received wisdom, often from religious authorities. Methods were questioned, interpretations of results were challenged, and scientists were attacked when science threatened to undermine religious and civil authority. In most circles, though, these sciences prevailed because the soundness of the methods, data, and interpretation withstood objective scrutiny. Social science has not yet achieved that status.

What accounts for the difficulty in achieving robustness in social science theory? Is it the very complexity of social behavior? Problems with the way data are collected, analyzed, and interpreted? Or is it that much of the subject matter of social science is at the heart of political and religious spheres of interest? Astronomy may have benefited from the fact that the power of political and religious authorities did not in fact reside in whether the earth was at the center of the universe. Biology faced a stronger challenge, but it, too, benefited when political and clerical leaders realized that the origins and evolution of species were not central to their sphere. Social science, though, must address topics that are at the heart of political and religious discourse.

Moreover, scientists are also human beings, behaving in social situations, as they practice science. A physicist who changes her position on, say, string theory may face social pressure for and against the change, but string theory does not lie at the heart of public policy debates. One's position on string theory is not likely to affect one's vote in the next election. But social scientists who study violence in inner cities, for example, are studying issues that do affect elections. To proclaim theories that question the wisdom of public policies related to such violence is to court opposition from supporters of those policies, and adulation from those who wish to change them. There is little reason to think that either side is particularly interested in science for science's sake.

The disciplines that seek to address social phenomena include experimental psychology, social psychology, sociology, cultural anthropology, cognitive science, medicine, evolutionary biology, and political science. The computational sciences are increasingly interested in

addressing social phenomena, and environmental sciences are also quite relevant in this arena. It is daunting to imagine theories and methods that could satisfactorily span these disciplines.

II. REVOLUTIONARY CONCEPTS

With respect to reproducibility, repeatability, and generalization of experiments, social scientists must accept that there is a problem and mount efforts to address it [5]. As other sciences matured, repeatability of results became expected, and lack of repeatability besmirched both the scientist who reported the study and the theory it supported. Social science must reach this point of maturation.

To reach this point, social science must develop a culture of sharing data, and agreement on methods of measurement and analysis. Repeatability of results cannot be expected when constructs are not defined the same way and measured the same way across studies. Results from studies that lack internal validity cannot be expected to have external validity, that is, to be generalizable beyond the specific conditions under which those results were obtained. Thus, long-term success in social science must address construct definition, measurement methods, and theoretical frameworks that span multiple academic disciplines.

A revolution in computational resources available to support social science is underway. “Big data” gives rise to the need for big platforms that support collection, maintenance, and sharing of social science data. Advances in machine-based text processing will produce methods to automatically scour the world’s literature for new findings, new methods, and new interpretation of vast repositories of social science data. Cognitive systems may well scrutinize published studies and identify the topic studied (even if called by different names in different studies), results and interpretation, and potential errors and biases in the study. This will allow for ongoing meta-analyses of prior studies and assimilation of multiple diverse data sets. For social scientists to understand these analyses, new visualization techniques are needed, and machine-generated interpretations expressed in natural language must accompany those visualizations

We can envision, then, a future in which social science studies are routinely conducted in the context of massive, ongoing collection of data about human behavior around the globe. These data sets will include everything from casual social media utterances to economic and policy decisions made by corporate and government leaders. One source of enabling technology is what is being called the Internet of Things (IoT): data from cyber physical objects such as mobile phones, automobiles, and home appliances will provide data about the behavior of people using those things. These data sets can be continuously updated. New hypotheses can be generated by scientists and by software systems, and competing theories can be subjected to ongoing tests as new data arrive.

A comparable situation emerged in meteorology as the community converged on the attributes to measure, the measurement methods, data representation conventions, and

protocols for sharing. Nowadays, a typical study in meteorology does not necessarily involve developing new measurement capability and collecting new data (although such studies do exist). Rather, a typical study might simply involve formulating a new hypothesis about causal mechanisms in weather patterns, and testing that hypothesis using massive data sets freely available across the community.

Perhaps social analytics will follow a similar pattern. Perhaps the globe will be instrumented with data collection capability for social phenomena the way that it is instrumented for local temperature, wind, and precipitation. These social data will be validated and loaded into accepted registries, and will immediately be used to update ongoing studies. New studies can be implemented in those registries, to test new hypotheses about causal mechanisms in human social behavior. These could be exciting times for social scientists.

Even more exciting is how these capabilities can positively impact the human condition – not just the advancement of science. These new capabilities can help us address social problems more effectively – not just measure them more reliably. Problems related to human health, standard of living, and subjective well being are intricately related to the phenomena studied by social scientists. In the developed democracies, re-election of incumbents is also affected by these phenomena. Politicians and business leaders alike will have a vested interest in the integrity of the social sciences and will therefore be more likely to keep them properly resourced.

III. THE ALTERNATIVE

The alternative is continuation of business as usual in the social sciences. Studies can continue to be conducted with college sophomores or other convenient samples of small size, and results will continue to lack robustness. Social science will continue to lack full respect in the broader scientific world, and support for social science will be driven by the usefulness of its finding to policy proponents.

Funding will not immediately cease, but it will not expand to the scale necessary for social science to succeed. Researchers in other disciplines will attempt to study social phenomena, and will attract funding and other forms of support in part due to the lack of acceptance of the established social sciences.

Most tragically, though, is that some of the core challenges of the human condition will continue to evolve without the benefits that rigorous science could potentially provide. Throughout the developing world, and in many population segments in the developed world, such problems as infant mortality, vulnerability to crime, malnutrition, unemployment, financial insecurity, and mental illness remain rampant. Vulnerable populations continue to be at higher risk in terms of health outcomes, economic outcomes, and social outcomes because of these problems.

Policy makers might well be willing to help alleviate those problems if only they could get guidance on the steps to take. Widespread adoption of vaccination for childhood

diseases occurred once medical science was able to identify and understand the disease and to develop effective methods of prevention. Until then, policy makers were divided on approaches to address such problems as polio. Once an effective vaccine was developed, policy fell in line, and those problems were greatly reduced. Similar advances are needed in the social sciences for the social problems that plague humanity across the globe. Until then, social science will continue to have a diminished place in the public forum.

In a special session on Next Generation Social Analytics, held as part of the HUSO 2016 conference in Barcelona, Spain [3], two papers are presented that discuss the challenges faced by, and payoffs expected from, the tools and methods that will facilitate conduct of social science research. Hutto [4] describes a vision of future social analytics rooted in large sample sizes, sophisticated measurement techniques, multiple epochs of data collection, and automated tools to extract data of interest. He uses a large data set compiled from multiple sources over a period of 40 years to illustrate how future studies can be conducted with disparate data sets and measurement techniques. McDermott, Nadolski, and Folds [5] describe a shared data experimentation platform that will transform how social science datasets are accrued, updated, and shared widely across a community of collaborating researchers. Such a platform will enable the transparency and open access that will enable confidence in data integrity, measurement methods, and analytic techniques.

REFERENCES

[1] Roberto Unger on what's wrong with social science today: <http://www.socialsciencespace.com/2014/01/roberto-mangabeira-unger-what-is-wrong-with-the-social-sciences-today/>

[2] "Estimating the reproducibility of psychological science". *Science*. **349** (6251): aac4716. August 28, 2015.

[3] HUSO 2016, The Second International Conference on Human and Social Analytics, November 13 - 17, 2016 - Barcelona, Spain, <http://www.iaia.org/conferences2016/HUSO16.html>

[4] C. Hutto, Blending Quantitative, Qualitative, Geospatial, and Temporal Data: Progressing Towards the Next Generation of Human Social Analytics, The Second International Conference on Human and Social Analytics November 13 - 17, 2016 – Barcelona (HUSO 2016), Spain, <http://www.iaia.org/conferences2016/HUSO16.html>

[5] T. McDermott, M. Nadolski, and D. Folds, System-Level Experimentation: Social Computing and Analytics for Theory Building and Evaluation, The Second International Conference on Human and Social Analytics November 13 - 17, 2016 - Barcelona, Spain, <http://www.iaia.org/conferences2016/HUSO16.html>