# Bigger Computational Social Science: Data, Theories, Models, and Simulations—Not Just Big Data

Claudio Cioffi-Revilla, Ph.D.
Center for Social Complexity and Department of Computational and Data Sciences
George Mason University
Research Hall, 4400 University Drive, Fairfax, Virginia 22030, USA
Email: ccioffi at gmu dot edu
May 24, 2016

***Abstract*—Computational social science (CSS) is an interdisciplinary field of social science that integrates individual social science disciplines. Its purpose is to advance scientific understanding of social phenomena through the medium of computing, which is used both as a paradigm and a methodological tool. Recently, restrictive versions of CSS have been proposed, based on "big data" now available from social media and other sources and progress in algorithms from computer science, while eschewing theory, models, or computational simulations—all three major parts of CSS. This paper argues for a comprehensive and balanced CSS that is paradigmatically guided by theory, enriched by analytical models, and enabled by computer simulations, all three drawing on data, be it big or small.**[1]

*It was six men of Indostan*
*To learning much inclined,*
*Who went to see the Elephant,*
*(Though all of them were blind)*
*That each by observation*
*Might satisfy his mind.*

J. G. Saxe (transl.), *The Blind Men and the Elephant*

## I. Introduction

Big data is valuable for expanding the frontiers of social science. It can be said that big social data or big data about social phenomena is already beginning to play a transformative role as a driver of growth in social knowledge. It is not surprising that big data should also be part of computational social science (CSS). However, CSS is an interdisciplinary field of social science that integrates individual social science disciplines. Its purpose is to advance scientific understanding of society and social dynamics using the computational paradigm of complex adaptive systems, as pioneered by Herbert A. Simon and his contemporaries, and to make advanced use of computation in *all* its functions. Recently, a highly restrictive version of CSS has been proposed, based on the new "big data" from social media and other sources, using available progress in algorithms from computer science, largely ignoring rigorous theory, models, or computational simulations—all three of which were championed by Simon. This paper argues for a comprehensive and balanced CSS, one paradigmatically guided by theory, enriched by analytical models, and enabled by computer simulations, all three drawing on data, be it big or small (Kitchin and Lauriault 2014).

This brief paper would be entirely unnecessary were it not for the fact that several recent publications by a number of accomplished scholars have presented a description of CSS focused entirely on the new big data (a sort of "Big Data CSS"; e.g., Lazer et al. 2009; Alvarez 2016). By well-established scientific and epistemological standards, such a view is incomplete and biased as a viable description of CSS, because it excludes other constituent components of the emerging field, such as theories, models, and simulations.

## II. Two Metaphors of the Situation

The current situation with Big Data CSS is reminiscent of two metaphors in which, by analogy, narrow perspectives erroneously claim to represent a much bigger and richer reality. In the first metaphor, a group of scientists gathers for a pasta meal consisting of *pappardelle alla bolognese* and some hearty *Amarone* wine (adapted from McCain and Segal 1988). The physicist sees how heat applied to salted water makes the latter boil sooner when cooking the pasta, and how torque exerted by his fork gathers each clump of pasta while wondering whether surface tension may attract more sauce. The biochemist understands how chewing the pasta begins to break down the grain flour glutens, assisted by the sauce's own acidity (pH $\approx 4.2$). The management scientist, a former systems engineer, thinks about the complex supply chains that make possible the delicious meal, providing pasta, San Marzano tomatoes, olive oil, ground meats, and all other necessary ingredients

[1]WORKING PAPER – COMMENTS ARE WELCOME. This paper is dedicated to the memory of two CSS pioneers: Herbert A. Simon, on the centennial of his birth (June 15, 1916), and John von Neumann, on the semi-centennial of his *Theory of Self-Reproducing Automata* (1966). Preferred citation: Claudio Cioffi-Revilla, "Bigger Computational Social Science: Data, Theories, Models, and Simulations—Not Just Big Data." Paper presented at the 8th International ACM Web Science Conference (WebSci'16), Hanover, Germany, May 22–25, 2016.


Electronic copy available at: http://ssrn.com/abstract=2784278

for a proper bolognese sauce, not to mention the cheese and white béchamel sauce! The sociologist enjoys his pasta while observing the table manners of his colleagues, two of which assist their fork with a soup spoon while another cuts her pasta instead of twisting it unbroken, indicative of their likely lower social status. The chef knows all these (and other) perspectives, so he has the most comprehensive and, therefore, most realistic understanding of the whole situation.

The other is the proverbial parable of the blind men and the elephant. Each blind man examines the elephant in its various parts, including tusks, feet, ears, back, tail, and other parts. They then compare notes and conclude that they are in disagreement, based on their individual perception. But the elephant's sighted owner would have had a realistic understanding of the whole animal, based on the power of inclusive integration.[2]

CSS is a large field consisting of several areas of investigation that offer immense and exciting domains and tools for research, such as algorithmic information extraction from big and small data sets, network models, social complexity, and various computer simulation traditions. Each of these can be pursued individually or in combination, depending on the goals of investigation. For one field to claim a monopoly on the whole field is like one of the actors in the two metaphors claiming that theirs is the whole field, while ignoring other perspectives. CSS is immensely exciting in proportion to the variety and power of its algorithms, data, theories, models, and simulations, especially when leveraged in combination and through synergies.

## III. Main Argument: "Big CSS"

There are three classes of epistemological reasons why CSS requires—and in fact demands inclusion of—theories, models, simulations, and other scientific constructs besides algorithms for big data. These reasons can be summarized in terms of three simple arguments stated in the form of claims, or even testable (i.e., refutable and verifiable) hypotheses:

1) CSS is a science.
2) CSS is a social science.
3) CSS is computational.

Each of these is documented in volumes (e.g, Bankes et al. 2002; Bennato 2015; Castelfranchi 2001; Cioffi 2009, 2010, 2014; Conte et al. 2012; Epstein 2006; Gilbert 2010; Hedström and Bearman 2009; Kuznar 2006; Manzo 2014; Squazzoni 2008; Torrens 2010; Trobia 2001), so what follows is merely a summary intended for reference.

### A. *CSS is a science*

CSS is part of science, as opposed to the fine arts or humanities. This basic fact indicates that the epistemology of CSS has to do with systematic, testable, and reproducible descriptions, explanations, and predictions (or forecasting)—although the third class is not strictly necessary for CSS to qualify as a science.[3]

Description requires observation and measurement, the results of which are often formalized in terms of mathematically stated laws that describe how phenomena occur. Duverger's Law of Party Systems, Zipf's Rank-Size Law, Pareto's Law on Inequality, and Richardson's Law of War Magnitudes, among many others, are examples. Although not all scientific laws are stated in mathematical form, those that are present many advantages, not the least of which is the power to generate additional logically valid and empirically testable deductions for expanding frontiers of knowledge.

Scientific laws that are stated mathematically are commonly referred to as models, although the term "model" includes many other objects, not all of which need be mathematical—analogous to the term "mechanism" in analytical social science and CSS. For example, a flowchart, a Gantt chart, and any UML diagram (class, sequence, state, use case, and others) are models. Models provide ways of describing phenomena, sometimes crossing the boundary between description and explanation.

Explanation in science means providing the causal mechanisms that account for observed phenomena, such as events and other occurrences, features, and laws. Mechanisms can be deterministic or probabilistic, depending on their causality, where the "or" refers to the Boolean logic inclusive OR, which is known as "and/or" in common language. Laws require theoretical explanation, because by themselves they only describe; they do not explain why a given pattern holds as a law. A scientific explanation is normally in the form of a process composed of one or more causal mechanisms.

Simulations are used in every field of science where there is need to analyze high-dimensionality models or theories, because traditional closed form solutions are impossible or undesirable for a variety of reasons. High dimensionality is a sufficient condition for using simulations, which enable the deduction of conclusions by running a simulation.

In all science, therefore, laws describe and theories explain—a synthesis due to S. Toulmin (1967). Laws and theories are constitutive or defining features of science, not optional activities or exotic embellishment created for esthetic purposes (although elegance is a desirable attribute of theories; Lave and March 1975). Simulations are necessary only when scientific analysis requires many variables or parameters that are relevant to the phenomenon under investigation, as are all complex systems and many complicated ones. This is the first reason why theories, models, and simulations are part of and necessary for CSS.

### B. *CSS is a social science*

The universe of social science includes many large and well-established disciplines (the "big five" are anthropology,

[2]See www.wordfocus.com/word-act-blindmen.html for John G. Saxe's English translation of the parable of *The Blind Men and the Elephant*. In the same website it is observed that "the story is also used to teach tolerance for other cultures. We only 'see' the culture in which we are immersed."

[3]The argument that CSS is part of science is valid, regardless of whether prediction or forecasting are viewed as features of science.


Electronic copy available at: http://ssrn.com/abstract=2784278

economics, political science, social psychology, and sociology) and interdisciplinary fields (e.g., human geography, management science, linguistics, communication science, social science history, among others). At the US National Science Foundation, the Directorate for Social, Behavioral, and Economic Sciences (SBE) includes programs in all of these domains, plus others, including some that link social science to other areas of biological, physical, and engineering sciences.

CSS is a social science. This idea is wholly similar to the notion that computational astronomy, computational biology, and computational linguistics are part of astronomy, biology, and linguistics, respectively. Any computational X-science is, by definition, part of X-science.

The social sciences include a large number of theories, models, and simulations, more in some areas and less in others, but *none of the social sciences is devoid of theories, models, and simulations for discovery and understanding of social phenomena*.[4] Social theories are used in CSS to inform the causal mechanisms and processes under investigation. Conversely, CSS contributes to social science by formulating new theories that explain social phenomena.

Social simulations are used when models and theories contain too many variables for mathematical or analytical approaches leading to closed form solutions. This situation is common, given the intrinsic complexity of even (seemingly) simple social phenomena (e.g., aggregating votes, small group behavior, decision-making under risk).

### *C. CSS is computational*

Social science became computational after centuries of having become historical, statistical, and mathematical—along with the natural and engineering sciences—soon after the invention of automatic computing; i.e., since the early 1960s (e.g., Guetzkow 1963; Simon 1969 [first edition of Simon 1996, cited]; Deutsch 1963; Messick 1963), about a decade after von Neumanns (1951) pioneering Theory of Automata. The fact that the term "computational" was not used from the very first day when computers were used to conduct social research does not detract from the fundamental claim.

As stated elsewhere, "Field Theory (Lewin 1952), Functionalist Theory (Radcliffe-Brown 1952), Conflict Theory (Richardson 1952a, 1952b), the Theory of Groups (Simon 1952), Political Systems Theory (Easton 1953), as well as Decision-making Theory (Allais 1953), among others, required new formalisms that could treat conceptual and theoretical complexity of human and social dynamics, beyond what could be accomplished through systems of mathematical equations solved in closed form" (Cioffi 2014: 24). Each of these social theories and related models, as well as numerous others, underwent computational implementation and simulation became a new methodology (Benson 1961; Borko 1962), before object-orientation was introduced into social research through R, Java, and other programming languages.

Earlier it was stated that a computational approach—including but not limited to simulation modeling methods—is required when encountering problems of high-dimensionality, as with many social phenomena. However, the computational approach in social science has a dual nature, encompassing a theory as well as a tool. This is because CSS is both a field of science enabled by computing, as well as a field of social science informed by a computational paradigm.

The methodological aspect of CSS is well-known and obviously due to the fact that computing plays a central role. Less appreciated or understood, but far deeper from a scientific perspective, is the computational paradigm of CSS introduced by Herbert A. Simon (1996) and others (Augier and March 2005; Batty 2006; Cioffi 2014; Holland 1975, 1995; Miller and Page 2007). This aspect of CSS relies on the interdisciplinary theory of complex adaptive systems and the role played by information processing at all scales of human and societal behavior. This computational paradigm also includes a specific Theory of Artificial Systems that explains how and why social complexity originates and evolves. Recent developments of the same computational and information-processing paradigm include the Canonical Theory of Social Complexity (Cioffi 2005, 2014: 214–220), which aims to explain the genesis and evolution of coupled and complex systems-of-systems that include human, artificial, and natural entities.

Again, big data also plays a major role at this computational level, because large amounts of information are necessary for numerous reasons that range from creating new theories and simulations to testing hypotheses and models of social systems and processes.

## IV. Conclusion

Computational social science is a science, a social science, and a field similar to other computational sciences where computing plays a dual and defining role as both paradigm and instrument. The former regards the fundamental role of information-processing for describing and understanding social phenomena, in this case through human decision-making at the micro-level that generates societal phenomena at the macro-level. The latter regards the role of the computer as an enabling instrument allowing us to reach and investigate frontiers of knowledge far beyond what is possible through historical, statistical, and mathematical methods. Both aspects define the character of CSS.

Today, theories, models, and data (both "small" [i.e., traditional] and "big" [more recent]) comprise the complex landscape and rich ecology of CSS. The only difference in terms of big data is that its sheer size and rate of growth (first and second derivatives, respectively) present major and exciting challenges to the other components. Theories and models are challenged and enabled by big data, and vice versa. Computational social simulations use and also create new demands for big data.

[4] If anything, it could be argued that a fundamental problem in social science is the proliferation of unviable theories lacking in testability or falsifiability, as well as the survival of demonstrably inferior or false theories. CSS can play a useful role in this area by demonstrating or exposing logical or empirical errors in putative social theories.

There is good reason to think that "the science of the twenty-first century will be computational" (Peter J. Denning). Such a big CSS perspective includes big data, algorithms, theories, models, and simulations, among other scientific constructs and instruments. Denning's insightful prediction is arguably becoming true of social science, as for most other domains of science. In the future, social science will prosper if it adopts the computational approach and, even more so, leverages more formal approaches from mathematics, just as during the twentieth century it learned how to apply statistics. The synergy of current statistical approaches, long-range historical and comparative perspectives, enhanced variety of mathematical approaches (beyond differential equations and game theory), and the full spectrum of computational approaches cannot but generate an explosive abundance of new scientific discoveries and deeper understanding of social phenomena, both simple and complex.

*So oft in theologic wars,*
*The disputants, I ween,*
*Rail on in utter ignorance*
*Of what each other mean,*
*And prate about an Elephant*
*Not one of them has seen!*

J. G. Saxe (transl.), *The Blind Men and the Elephant*

Computational social scientists should work together to describe, explain, and scientifically understand the whole elephant of social science.

## Acknowledgment

This invited paper was prepared for presentation at the panel on "Computational Social Sciences: A Bricolage of Approaches," 8th International ACM Web Science Conference (WebSci'16), Hanover, Germany, May 22–25, 2016, chaired by Paolo Parigi. The author would like to thank panel chair Paolo Parigi for the invitation to write and present this paper, and Robert Axtell, Joshua Epstein, Nigel Gilbert, Dirk Helbing, David Masad, J. Daniel Rogers, Larry Kuznar, Chris Rouly, Flaminio Squazzoni, Qing Tang, and Klaus G. Troitzsch for comments. This paper was funded by NSF CDI Type II grant no. IIS-1125171 and by the Center for Social Complexity at George Mason University. The author is solely responsible for the views and opinions expressed in this paper.

## References

Allais, Maurice. (1953). Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'ecole Americaine [Rational human behavior under risk: critique of postulates and axioms of the American school]. Econometrica, 21, 503–546.

Alvarez, R. Michael (2015). Computational Social Science: Discovery and Prediction. Cambridge University Press.

Augier, Mie, & March, James G. (Eds.). (2004). Models of a Man: Essays in Memory of Herbert A. Simon. Cambridge, MA: The MIT Press.

Batty, Michael. (2006). Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. Cambridge, MA: The MIT Press.

Bankes, Steven C., Lempert, Robert J., & Popper, Steven W. (2002). Making Computational Social Science Effective: Epistemology, Methodology, and Technology. Social Science Computer Review, 20(4), 377–388.

Bennato, Davide. (2015). Il computer come macroscopio [The Computer as a Macroscope]. Milan, Italy: Franco Angeli.

Benson, Oliver. (1961). A Simple Diplomatic Game. In James N. Rosenau (Ed.), International Politics and Foreign Policy (pp. 504–511). New York: Free Press.

Borko, Harold (Ed.) (1962). Computer Applications in the Behavioral Sciences. Englewood Cliffs, NJ: Prentice-Hall.

Castelfranchi, Cristiano. (2001). The theory of social functions: challenges for computational social science and multi-agent learning. Journal of Cognitive Systems Research, 2, 5–38.

Cioffi-Revilla, Claudio. (2005). A Canonical Theory of Origins and Development of Social Complexity. Journal of Mathematical Sociology, 29(AprilJune), 133–153.

Cioffi-Revilla, Claudio. (2009). On Social Complexity: A Manifesto for Computational Social Science. Presidential Address. Annual Conference of the North American Association of Computational Social and Organizational Sciences, Phoenix, Arizona.

Cioffi-Revilla, Claudio. (2010). Computational Social Science. Wiley Interdisciplinary Reviews (WIREs): Computational Statistics, 2(3), 259–271.

Cioffi-Revilla, Claudio (2014). Introduction to Computational Social Science: Principles and Applications. London and Heidelberg: Springer.

Conte, R., Gilbert, G. Nigel, Bonelli, G., Cioffi-Revilla, C., Deffaunt, G., Kertesz, J., . . . Helbig, D. (2012). Manifesto of Computational Social Science. European Physical Journal Special Topics, 214, 325–346.

Deutsch, Karl W. (1963). The Nerves of Government. New York: Free Press.

Easton, David. (1953). The Political System: An Inquiry into the Study of Political Science. New York: Alfred A. Knopf, Inc.

Epstein, Joshua. (2006). Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton, NJ: Princeton University Press.

Gilbert, Nigel (Ed.) (2010). Computational Social Science. Los Angeles: Sage.

Guetzkow, Harold, Alger, Chadwick F., Brody, Richard A., Noel, Robert C., & Snyder, Richard C. (Eds.). (1963). Simulation in International Relations: Developments for Research and Teaching. Englewood Cliffs, NJ: Prentice-Hall.

Hedström, Peter, & Bearman, Peter (Eds.). (2009). The Oxford Handbook of Analytical Sociology. Oxford University Press.

Holland, John H. (1975). Adaptation in Natural and Artificial Systems. Ann Arbor, MI: University of Michigan Press.

Holland, John H. (1995). Hidden Order: How Adaptation Builds Complexity. Reading, MA: Addison-Wesley.

Kitchin, Rob, & Lauriault, Tracey P. (2014). Small data in the era of big data. GeoJournal, 80(4), 463–475.

Kuznar, Lawrence A. (2006). High-Fidelity Computational Social Science in Anthropology. Social Science Computer Review, 24(1), 15–29.

Lave, Charles A., & March, James G. (1975). An Introduction to Models in the Social Sciences. New York: Harper and Row.

Lazer, David, Pentland, Alex, Adamic, Lada, Aral, Sinan, Barabasi, Albert-Laszlo, Brewer, Devon, . . . Alstyne, Marshall Van. (2009). Computational Social Science. Science, 323(5915), 721–723.

Lewin, Kurt. (1952). Field Theory in Social Science: Selected Theoretical Papers. Chicago and London: University of Chicago Press.

Manzo, Gianluca (Ed.) (2014). Analytical Sociology: Actions and Networks. New York: Wiley.

Masad, David. 2014. Computational social science, data, and complexity. http://davidmasad.com/blog/css-data-complexity/

McCain, Garvin, & Segal, Erwin M. (1988). The Game of Science (5th edition ed.). Pacific Grove, CA: Brooks/Cole Publishing Co.

Messick, Samuel J. (1963). Computer Simulation of Personality: Frontier of Psychological Theory. American Psychological Association.

Miller, John H., & Page, Scott E. (2007). Complex Adaptive Systems: An Introduction to Computational Models of Social Life. Princeton and Oxford: Princeton University Press.

Radcliffe-Browne, A. R. (1952). Structure and Function in Primitive Society. Glencoe, Illinois: Free Press.

Richardson, Lewis Fry. (1952). Is it possible to prove any general statements about historical fact? British Journal of Sociology, 3(1), 7784.

Richardson, Lewis Fry. (1952). Contiguity and deadly quarrels: The local pacifying influence. Journal of the Royal Statistical Society, Series A, 115(Part II), 219-231.

Simon, Herbert A. (1952). A formal theory of interaction in social groups. American Sociological Review, 17(2), 202-212.

Simon, Herbert. (1957). Models of Man. New York: John Wiley and Sons.

Simon, Herbert A. (1965). The architecture of complexity. Proceedings of the American Philosophical Society, 106, 467-482.

Simon, Herbert A. (1996). The Sciences of the Artificial (3rd ed.). Cambridge, MA: MIT Press.

Squazzoni, Flaminio. (2008). A (computational) social science perspective on societal transitions. Computational and Mathematical Organization Theory, 14(4), 266–282.

Toulmin, Stephen. (1967). The Discovery of Time. London: Penguin.

Trobia, Alberto. (2001). La Sociologia Come Scienza Rigorosa: Modelli simulativi, intelligenza collettiva, forme del mutamento [Sociology as a Rigorous Science: Simulation Models, Collective Inteligence, and Patterns of Change]. Milan, Italy: Franco Angeli.

von Neumann, John. (1951). The General and Logical Theory of Automata. In A. L. Jeffress (Ed.), Cerebral Mechanisms in Behavior (pp. 141). New York: John Wiley.

von Neumann, John. (1966). Theory of Self-Reproducing Automata. Urbana, Illinois: University of Illinois Press.