DGOF

NEUE SCHRIFTEN ZUR ONLINE-FORSCHUNG 15

COMPUTATIONAL SOCIAL SCIENCE IN THE AGE OF BIG DATA

CONCEPTS, METHODOLOGIES, TOOLS, AND APPLICATIONS

CATHLEEN M. STUETZER / MARTIN WELKER / MARC EGGER (HRSG.)

HERBERT VON HALEM VERLAG

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists the publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Cathleen M. Stuetzer / Martin Welker / Marc Egger (Eds.)
Computational Social Science in the Age of Big Data.
Concepts, Methodologies, Tools, and Applications
Neue Schriften zur Online-Forschung, Band 15
Köln: Halem, 2018

ISSN 1865-2638

ISBN (Print): 978-3-86962-267-5
ISBN (PDF): 978-3-86962-268-2

http://www.halem-verlag.de
info@halem-verlag.de

TYPESETTING: Herbert von Halem Verlag
EDITOR: Imke Hirschmann, Köln
PRINT: docupoint GmbH, Magdeburg
COVERDESIGN: Claudia Ott, Grafischer Entwurf
Copyright Lexicon ©1992 by The Enschedé Font Foundry.
Lexicon® is a Registered Trademark of The Enschedé Font Foundry.

NEUE SCHRIFTEN ZUR ONLINE-FORSCHUNG

Cathleen M. Stuetzer / Martin Welker / Marc Egger (Eds.)

Computational Social Science in the Age of Big Data

Concepts, Methodologies, Tools, and Applications

HERBERT VON HALEM VERLAG

Inhalt

III. CASE STUDIES

PREFACE

CATHLEEN M. STUETZER / MARTIN WELKER /
MARC EGGER

Big Data Analytics: Obstacles and Opportunities for Social Science

Digitalization has already permeated all areas of life and is omnipresent – not only in our everyday life, but also in politics, business and especially in science. Industry is now talking about wearables, the Internet of Things and industry 4.0 when it comes to promoting digitalization and digital networking within the production cycles. Over the past years, the velocity of technolgical advancements has tremendously increased that the exploitation of the so-called *digital footprints* – which we leave permanently in our everyday life – becomes the subject of social science research.

In the last 15 years with the emergence of social media platforms, we observe a new phase of data revolution (LAZER et al. 2009; ALVAREZ 2016). With the massive increase in data production Big Data is more and more discussed as *socio-technological* phenomenon (BOYD/CRAWFORD 2012; LAZER et al. 2009; ALVAREZ 2016). With the help of computational techniques the collection and extraction of data seem often easy to obtain and cheap (KING 2016). Thus, in social science we notice a computational turn in research, and that is mostly associated with high expectations on scientists (BOYD/CRAWFORD 2012). But how can we benefit from analyzing big (social) data? How can we handle the new data? Which analytical approaches, techniques, and instruments are actually used and discussed? Which skills are needed in that upcoming field? What should we know about ethical and privacy issues? And what are the consequences for theoretical considerations?

Although research on Big Data has a long tradition, only with the emergence of online communities and social media platforms in 1990s' Big Data arise as socio-technological phenomenon (BOYD/ELLISON 2007). Actually, Big Data is a term which influences all fields of (applied) research, and, according to Boyd and Crawford (2012), »Big Data not only refers to very large data sets and the tool and procedures used to manipulate and analyze them, but also to a computational turn in thought and research« (p. 665). But what is the meaning of this computational turn in thought and research?

Traditionally, researchers are focused on answering research questions regarding a special target group. In the context of Computational Social Science (CSS), data is often captured first. Thus, the process of exploration of massive social data plays an important role to identify a focus group and generate hypothesis as well as research questions after that. It seems to be a fruitful opportunity to get new insights of social phenomena.

New perspectives on data analytics bring a lot of obstacles and opportunities for researchers. First of all, a new paradigm raises high expectations. According to Alvarez (2016) we »examine the social world in new ways« (p. 4) and bring Big Data to life. Second, we have always sought answers to the question on how the (social) world is constructed, and we tend to reflect new approaches critically. Third, with the emergence of social technology we need suitable infrastructures – not only technological and methodological but also socio-cultural. The opportunities are obvious. CSS based on theoretical approaches which explain social world as a connected world in different stages. From this point of view, we get insights about social processes related to communication, interaction, and social relations on the (social) web at different levels. By analyzing a new type of data and by using new techniques, we can extract new types of knowledge (LAZER et al. 2009). Continuing to this, we identify Big Data as a »social construct« to handle (social) phenomena within the connected world.

CSS as methodological approach offers systematized ways by using new techniques for collecting, extracting, and analyzing large-scale (social) data in (applied) research. CSS enables to track digital footprints and stream »social« traces with computational methods (LAZER et al. 2009; ALVAREZ 2016). New analytical approaches behind CSS allow to explore social behavior and the dynamics of causal correlations. CSS highlights opportunities in representing projections of the social world by analyzing digital traces people left behind (CIOFFI-REVILLA 2014). Most social traces are left using the Internet – but social science methods focused on human traces are older than the web (e.g. JAHODA/LAZARSFELD/ZEISEL 2015 [1933]). Nevertheless, nowadays we have a set of new possibilities reaching far beyond the methods and instruments already applied in the 1930s. Technologically driven changing communication habits change the circumstances under which social research can sensibly be conducted in contemporary society. Therefore, social scientists ask at least three intrepid questions when it comes to the field of digital methods and algorithms:

1. Does the emerging research field of Computational Social Science require a new methodology?
2. Does it need new, conceivably additional, scientific quality criteria? So, does Computational Social Science require new or modified methods within current methodology or is there a demand just for better performance in practice?
3. And if we need new or modified methods, what could it be, which new or improved computational methods are prerequisite to the field? How is an optimal integration of computational methods achievable?

The classical research procedure now shifts towards data collection, data processing, and data storage. Extremely, the course is turned upside down: assumptions and hypotheses are discussed at last, after outcomes are generated with theories totally eliminated. A new methodology could reflect this approach at a meta-level critically. Secondly, this new empirical field is strongly based on computer science and its subdisciplines. But humanities and social science have another focus while computer science is often focused on the optimization of machine processes. The discussion of algorithms for research due to methodological and ethical problems mirrors these different rationalities.

Under those circumstances, does Computational Social Science expect an original set of theories or a theoretical framework? In other words, does CSS require a theoretical basis that links all the diverse works and studies and serves as a common draw of the field? And if yes, what framework could we use? A theoretical framework would be desirable but perhaps not necessary. As Schroeder and Taylor (2015) and Stegbauer (2009) have shown, Big Data studies on Wikipedia are extremely heterogeneous. The Big Data studies on Wikipedia had no common perspective but a common goal: to be able to answer research and professional questions by applying Big Data. In this context, for example, innovation theories possibly are most suitable, also approaches of common understanding, while theories of action may be less appropriate. But how do perform the new field in practice? Does it meet the accuracy and precision scientists and society need?

Accordingly, large multinational telecommunication companies and online service providers install their own departments of data, thus making science proprietary instead of publicly debatable and verifiable. The digital market leaders like Apple, Google, Amazon, Facebook, Microsoft, etc. set up their own research basis, storing data and using their platforms as massive research tools. But what is the impact of this trend on academic social science? What are the consequences?

Nevertheless, the potentials of introducing CSS as research paradigm in social science become visible. As the learnings of Online-Research and of 20 years of experience in validating new methods and instruments show us, it seems that CSS comes with a certain momentum to justify a new theoretical and methodological basis. The use of new ways of (social) data collection, extraction and exploration open up completely new fields of activity. We are at the beginning of a new era in applying new research theories, methods, and applications in (applied) social science.

With this book we want to initiate a discourse – theoretical as well as analytically – about the upcoming field of research. We present selected contributions to demonstrate the computational turn in social research as well as the relevance for the applied market research. Our selected contributions in this book underpin that CSS is a young and highly interdisciplinary field of research that primarily aims to generate complex data to usable information. The exploration of massive data is not only interesting for computer scientists, physicians, meteorologists, and/or business economist but also for psychologists, social scientists, communication experts, and political scientists. On the one hand, it is attempted to gain insights into social phenomena from process-generated data – on the other hand, new methodological approaches are used to answering questions about (social) impact mechanisms. This field of research is driven by the primary research mission to contribute further developments of evidence-based behavioral and impact research.

Within the first chapter »Epistemological Perspectives«, the authors open the discussion about the relevance for CSS as research field. They highlight obstacles and opportunities of Big Data analytics in social science. They explore e.g. the bridge between social and computational science, demonstrate theoretical approaches, discuss applicable methodologies, point out fields of activities, and illustrate limitations and restrictions in that field.

The second chapter »Data, Methods, and Instruments« deals with analytical and methodological approaches in the field of CSS. The authors present different techniques and instruments to handle massive (online) data. They introduce us in e.g. (automated) data collection, tracking methods for analyzing social behavior, storage of large-scale (social) datasets, tools for collecting and storing historical financial data, potential benefits of using hierarchical as well as dynamic clustering approaches as well as information mapping and data visualization in the context of Big Data analytics.

The third chapter »Case Studies« puts theory into practice. The authors demonstrate the wide range of application on css theories, methods, and instruments. The studies explore the use of Big Data by examining e.g. human relationships and kinship, migration routes, and dynamics of protest, and the relevance for affective computing.

The fourth chapter »Tutorial Section« aims to introduce us in practical application of Computational Social Science. The authors demonstrate hands-on manuals in the field of learning analytics as well as social media monitoring.

References

ALVAREZ, R. (ed.): *Computational Social Science: Discovery and Prediction* (Analytical Methods for Social Research). Cambridge [Cambridge University Press] 2016

BOYD, D.; K. CRAWFORD: Critical Questions for Big Data. In: *Information, Communication & Society,* 15(5), 2012, pp. 662-679

BOYD, D. M.; N. B. ELLISON: Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication*, 13(1), 2007, pp. 210-230

CIOFFI-REVILLA, C.: *Introduction to Computational Social Science. Principles and Applications*. London [Springer] 2014

CONTE, R.; N. GILBERT; G. BONELLI; C. CIOFFI-REVILLA; G. DEFFUANT; J. KERTESZ; V. LORETO; S. MOAT; J. P. NADAL; A. SANCHEZ; A. NOWAK; A. FLACHE; M. SAN MIGUEL; D. HELBING: Manifesto of Computational Social Science. In: *The European Physical Journal Special Topics*, 214, 2012, pp. 325-346

JAHODA, M.; P. F. LAZARSFELD; H. ZEISEL: *Die Arbeitslosen von Marienthal. Ein soziographischer Versucht*. 25. Auflage, Frankfurt/M., Leipzig [Suhrkamp] 2015 [1933]

KING, G.: Preface: Big Data Is Not About the Data! In: ALVAREZ, R. MICHAEL (ed.): *Computational Social Science: Discovery and Prediction*. Cambridge [Cambridge University Press] 2016

LAZER, D.; A. PENTLAND; L. ADAMIC; S. ARAL; A. BARABÁSI; D. BREWER; N. CHRISTAKIS; N. CONTRACTOR; J. FOWLER; M. GUTMANN; T. JEBARA; G. KING; M. MACY; D. ROY; M. VAN ALSTYNE: Computational Social Science. In: *Science*, 323(5915), 2009, 721-723

SCHROEDER, R.; L. TAYLOR: Big Data and Wikipedia Research: Social Science Knowledge Across Disciplinary Divides. In: *Information, Communication & Society*, 18(9), 2015, 1039-1056

STEGBAUER, C.: *Wikipedia: Das Rätsel der Kooperation*. Wiesbaden [VS Verlag für Sozialwissenschaften] 2009

I. EPISTEMOLOGICAL PERSPECTIVES

Brenda L. Berkelaar / Luis Francisco-Revilla

Motivation, Evidence, and Computation: A Research Framework for Expanding Computational Social Science Participation and Design

Computational social science has captured the imagination of scholars and the general public. A cross-cutting research field, computational social science inspires researchers to explore avenues of inquiry unlocked by technological advances and rapidly expanding datasets (Lazer et al. 2009; Watts 2013). Computational analysis gives researchers the means to study vast amounts of user information harvested by social media, mobile devices, and communication technology. The mundane uses and affordances of these tools offer rich sources of evidence for understanding individual and collective human behavior.

Nevertheless, computational social science has not yet realized its full potential. Computational social science will continue to develop as computing power continues to grow. A relatively nascent research area, computational social science initially gravitated towards a narrow subset of social science questions, conventionally tethered to particular methodological approaches and communities. Yet, computational social science can help address the broader landscape of questions conventional social science engages. This chapter discusses the opportunities that emerge when computational social science expands its perspective on social science and computation.

Computational social science can benefit from more broadly mining the rich depths of social science's methodological, theoretical, and conceptual imagination. This requires:

- Connecting computation with a broader set of research questions, methodological approaches, and research communities; and
- Rethinking, recombining, and developing the computational tools necessary to answer such questions.

The approach presented here provides one way of leveraging the promises computational social science offers to the meaningful progress of computational and conventional social science (see LAZER et al. 2009; WATTS 2013). By engaging in ongoing conversations about what computational social science *is* and what it *is for*, researchers can better answer questions about social phenomena while also enhancing computational approaches.

Therefore, the purpose of this chapter is twofold: to illustrate the value of further cultivating a mutually beneficial relationship between computational social science and conventional social science; and to provide a framework for people interested in initiating or developing computational social science projects – especially projects which might not seem computationally relevant at first glance. Computational social science can inform conventional social science and conventional social science can inform computation and computational social science because:

- Computational social science challenges conventional social science to evaluate, refine, and expand theories given computational requirements and the theoretical and empirical potential of big data analyses; and
- Social science challenges computational social science to devise methodologies and workflows that allow researchers to leverage large-scale data observations to explain complex social phenomena.

Although such language suggests conventional social science and computational social science are distinct fields, in practice their respective foci overlap. Further, considering potential intersections between computational and conventional social sciences highlights rich possibilities for intellectual development in social science and in computation.

This chapter is divided into three sections. Section 1 discusses the origins and the current state of computational social science. Section 2 presents a workflow for a computational social science research. We illustrate this workflow with an exemplar project. Section 3 presents a research framework for computational social science research based on lessons learned thus far. The framework encourages researchers to examine the assumptions and motivations driving their research questions, as well as the purposes and potential of computational social science for understanding human behavior. In brief, this chapter highlights the needs and generative possibilities afforded by expanding the methodological and theoretical traditions considered when seeking to understand complex social scientific phenomena computationally.

1. Situating Computational Social Science

Computational social science promises to accelerate discovery and insight into social phenomena. It does so by leveraging computational algorithms and by increased access to large-scale social datasets or »big data« (LAZER et al. 2009). In its first stages, quantitative reasoning strongly influenced computational social science, as have the research domains conventionally tethered to quantitative approaches. Quantitatively-tethered questions, problems, and approaches rely on applying enough computational power and appropriate algorithms to analyze very large social data numerically. For example, digital breadcrumbs from communication technologies help test propositions about how large social networks function with what effect (NEWMAN 2003); how social ties form (KOSSINETS/WATTS 2006); and whether and when people join groups (BACKSTROM et al. 2011). Plus, crowdsourced virtual labs have helped simplify, expand, and streamline experimental research on social behavior – especially network experimental designs which benefit from access to large groups of participants engaged in simultaneous activities (see WATTS 2013).

The tendency for computational social science to be inspired by conventional quantitative approaches makes sense. The social scientists, physical scientists, and practitioners drawn to computational social science are often inclined towards quantitative approaches because of their training, apprenticeships, and professional practice. In addition, many researchers involved in computational social science have limited training in or knowledge of social scientific theories or research traditions (WATTS 2013). Instead, they hail from disciplines like physics and computer science, drawn in part by the large networked datasets generated from the pervasive use of contemporary communication technologies. Socialization and reward structures further reinforce such inclinations. Plus, common understandings frame computation as the use of powerful »number-crunching« machines to analyze big datasets. Consequently, it becomes easy to imagine why many conventional social scientists feel computational social science is irrelevant to their research interests or is inaccessible given the time and resources needed to learn computational approaches. Such factors help explain why computational social science inclines towards extensions of quantitative methods, as well as the implicit, if not necessarily required, overarching goals, topics, and theoretical commitments often entangled with particular methodological approaches.

Yet, social science involves more than quantification. Even as quantitatively-inspired computational approaches continue to provide valuable contributions, many social phenomena present tough non-numerical challenges for computational analyses. The study of social phenomena often requires interpreting information implicit in multilevel, complex, and emergent data. Relevant and substantive questions about the social phenomenon may not practically or philosophically lend themselves to quantification. In addition to testing theory or producing generalizations, a rich body of social science focuses on interpreting the diversity of human experience, identifying the cultural significance of social trends, or giving voice to marginalized groups, often by studying a small set of cases or outliers in depth (RAGIN/AMOROSO 2010). Moreover, different social science communities tend to privilege certain meta-theoretical commitments or methodological preferences in pursuit of specific questions about human behavior. These diverse commitments have provided influential ways of understanding and influencing society theoretically, practically, and ethically.

Consequently, the primary influence of quantitatively-inspired research approaches brings entanglements with particular methods, theories, philosophy, and goals. Such entanglements can unnecessarily circumscribe the topics and questions posed by computational social science – often through habit, if not through intention. Unnecessary circumscription can hobble the development of computational approaches that could address the broader goals and topical range of conventional social science by limiting the problem sets and use cases used to develop computational approaches. Evidence of this circumscription can be seen in the limited cross-pollination between influential computational social science and conventional social science (e.g., conferences, publications; WATTS 2013). Such intellectual isolation is likely due to a variety of factors: limited knowledge of social science theories and influential outlets on the part of many computational social scientists; limited understanding or uncertainty about computation by conventional social scientists (WATTS 2013); the relatively independent development of computational social science (LAZER et al. 2009; WATTS

2013); and the topical, theoretical, and methodological entanglements reinforced by professional socialization, practice, and reward structures (LAWSON 1995; MCEVOY/RICHARDS 2006; RAGIN/AMAROSO 2010).

The tendency for methodological preferences and expertise to become unnecessarily entangled with topical and metatheoretical commitments remains a persistent critique of contemporary social science. For example, researchers who self-identify as »quantitative« often hold implicit assumptions that the purpose of research is generalizable pattern recognition and theory testing, whereas researchers who self-identify as »qualitative« tend to focus on goals related to identifying and interpreting contextual and cultural significance, or to giving voice to marginalized groups (RAGIN/AMAROSO 2010). Such emphases are differences in degree rather than in kind; however, the methodological, philosophical, and topical entanglements underlying these emphases are consequential for the development and contributions of computational social science.

Such entanglements limit the understanding of social phenomena (FAY/MOON 1977). When asked, most researchers agree that the research question determines the method. In practice, however, people often chose research questions based on the methods and tools that reflect their expertise, habits, or both. Reinforced by professional socialization, research biographies, and limited time, this qualitative-quantitative dualism persists (LAWSON 1995). Yet, such habits, if left unexamined, can unnecessarily circumscribe the potential of computational social science to help explain the contexts and contours of human behavior effectively and ethically. To avoid being distracted by the glittery promises of computation and big data, researchers need to reflectively consider the epistemologies and ethics underlying computational projects (BOYD/CRAWFORD 2012; KITCHIN 2014).

We could take in this argument in at least two directions. The first path compares and contrasts qualitative and quantitative approaches before showing the generative value of both through mixed method research. Such an approach is increasingly valued in conventional social science (e.g., CRESWELL/CLARK 2010; RAGIN 2014). The second path encourages and reinforces a broader conceptualization of what computation is, and what computation is for. This approach requires setting the persistent qualitative-quantitative binary distinction aside, along with the intuitive tendency to think of computational social science as a process akin to »really big statistics«. This second, less-developed approach is the focus of this essay. A broader definition of computation can expand the range of topics, philosophies, methods, and expertise that inform and leverage computation and computational social science; it can increase the vision of active participants in computational social science; it can encourage a broader range of people to engage in computational social science. In sum, a broader definition of computation can help researchers address a broader range of questions about human behavior.

The next section outlines a pragmatic definition of computational social science. This definition refocuses computational attention on a more diverse range of social science topics and questions. In reframing computation, our goal is to build on the achievements and to extend the focus of computational social science to better fulfill its promised potential. We invite researchers to consider computational social science in new ways – especially researchers who might not otherwise consider their research computationally relevant.

1.1 *Clarifying Assumptions*

It is necessary to clarify a few assumptions as we tread into often contentious areas of social science. First, we have no qualms with adopting and adapting conventional quantitative approaches for computational social science. Such approaches continue to prove generative and influential. Rather we assert that computational social science can be enhanced by also leveraging insight from the conventional habits, practices, and foci of social science more broadly, in particular researchers often labeled »qualitative«. So-called »qualitative« researchers are often overlooked by computational social science and often overlook computational social science, despite the demonstrated potential of qualitative scholarship to enhance discoveries and insights into social phenomena. Computational social science can benefit by drawing inspiration and insights from the full range of methodological approaches and philosophical perspectives that conventional social science employs in service of the rich range of research questions social science endeavors to answer. By reframing our understandings of what computational social science is and can do, we seek to encourage a more diverse range of scholars to participate in computational social science.

Second, we recognize that differentiating between conventional quantitative and qualitative approaches to social science creates an artificial dichotomy. This dichotomy often oversimplifies methodological distinctions and conflates theoretical commitments with analytic decisions. Yet such methodological distinctions continue to shape the everyday practices and productions of social science through socialization practices and reward structures (LAWSON 1995; RAGIN 2014). The everyday shorthand researchers use to implicitly or explicitly identify each other as qualitative or quantitative has substantive implications for the breadth and depth of social science insights gained – especially if a one set of approaches dominates the development and trajectory of crucial avenues of inquiry.

The persistent, if weakening, supposition that qualitative and quantitative methods tend to be incommensurate is unwarranted (BRYMAN 1992; BRYMAN/BELL 2003; CRESWELL/CLARK 2010; HAMMERSLEY 1992; JOHNSON/ONWUEGBUZIE 2004; LAYDER 1993; TEDDLIE/TASHAKKORI 2009). Scholars often conflate quantitative methods with more tangible, realist ontologies, with the goal of verifying or falsifying hypotheses. In contrast, qualitative methods are often conflated with more intangible realities and social constructionism, with more interpretive or critical goals for analysis (MCEVOY/RICHARDS 2006). Yet, methodology should not be conflated with a method's technical aspects, nor should one presuppose a specific ontology or epistemology is necessary to employ a particular method (MCEVOY/RICHARDS 2006). For example, researchers with realist or interpretive ontologies might employ ethnographic approaches (VAN MAANEN 2011). Feminist post-structuralist as well as post-positivist researchers might employ quantitative methods like hierarchical linear modeling or semantic equation modeling (LAWSON 1995).

This chapter is not an argument to ignore philosophy. Researchers should be aware of the epistemologies, ontologies, and axiologies entangled in different research approaches. Talk about big data and computational social science creates profound epistemological and ethical changes with the potential to become crystallized orthodoxies (BOYD/CRAWFORD 2012), especially if researchers and the people they serve assume »numbers can speak for themselves« (ANDERSON 2008). Rather, computational approaches – and the data and assumptions that fuel them – need reflective contextualization and critique to fulfill the ethical and efficacy goals of social science.

Relatedly, disciplines often define qualitative and quantitative research differently. For example, some researchers consider content analysis »qualitative« given the focus on identifying themes, whereas others consider content analysis »quantitative«, given the focus of coding on counting the prevalence of specific themes. Since computational social science typically involves multi-disciplinary collaboration, being clear about the implicit assumptions and meanings of terms becomes key to successful collaborations. As part of the reflective practice of effective collaboration, researchers would be wise to explain the evolution of research traditions and associated assumptions from their disciplinary perspectives. Understanding the distinct, if overlapping, research practices and emphases of conventional qualitative and quantitative research approaches helps highlight potential challenges and moral dilemmas even as it expands possibilities for innovative research that better realizes the potential of computational social science.

Finally, the research framework we propose is neither the avenue, nor the only avenue, for rigorous, generative social science. Nor do we assert that all social science should be computational. Insights into human behavior have been enriched by the diverse range of methodological, theoretical, philosophical, and instrumental goals fueling social science. Rather, in offering an expansive, yet pragmatic, way of thinking about computation, we provide a generative way of thinking about the diversity of social phenomena. Our intent is to help people realize the potential capacity of computational social science to inform social science and computation in ethical, practical, and theoretical ways. Our intent is to encourage computational newcomers and experts with diverse backgrounds, training, and perspectives to initiate, develop, and collaborate on computational social science projects that address the range of substantive questions conventional social scientists ask. Thus, we focus less on how qualitative and quantitative perspectives differ or have differed. Instead, we uncouple qualitative and quantitative research from singular theoretical entanglements to refocus on a shared curiosity about social behavior. This uncoupling allows us to consider what each of these broad research traditions can offer the development of computational social science specifically, and conventional social science broadly.

1.2 *Defining Computational Social Science*

Defining computational social science is challenging. The term is a contested, moving target within a relatively nascent and rapidly evolving research area. Yet definitions matter. Definitions influence what is considered possible and desirable (PHILLIPS/OSWICK 2012). Implicit and explicit across varied definitions of computational social science is the idea that computational social science is inspired by, and takes advantage of, the growing breadth, depth, and amount of data available about human life (LAZER et al. 2009). Moving beyond a shared emphasis on big data, we take a pragmatic approach to defining computational social science – the domain where social science and computation intersect. Pragmatism focuses less on identifying the singular right or true definition and more on asking: What difference does it make if this were true (KAPLAN 1964)? That is, we ask what is the difference if computational social science involves more than organizing and calculating big data with machines as often defined? By reconsidering common understandings of what social science and computation are, we can expand understandings of what computational social science can do, could do, should do, and is likely to do.

As a major category of intellectual thought, *social science* focuses on asking questions about human behavior: Why do humans act the way they do and how do society and culture shape human life individually and collectively (ECONOMIC AND SOCIAL RESEARCH COUNCIL 2017)? Social science is rooted in the ancient Greeks' focus on the nature of humanity, morality, and society. Although recognized as a distinct discipline in the 19th century, debates continue over which disciplines and which topics and methods constitute »social science«. Rather than debating the exact boundaries of social science, we focus on where social science is anchored: Social science inquiry focuses on questions of human behavior – in particular, how individual behaviors influence collective, social life. Such a definition allows researchers from diverse backgrounds and foci to come together to provide insight into the complexities of individual and collective human behavior. Such a definition recognizes that insight into the complexities of human behavior is rarely situated within one particular discipline, field, or specialty; nor is it informed exclusively by particular methodological or metaphysical commitments. Broadly speaking, curiosity about human behavior drives social science research and, by extension, computational social science.

So then, what makes social science *computational*? To state the seemingly obvious: Computational social science involves computation. At first glance, such a statement seems self-explanatory. It is not. Computation itself is an idea in flux (DENNING 2010; HORSWILL 2012). Colloquially, people often equate computation with information technology: People view computation as the work done by computers as machines. In this sense, computation describes what digital machines do. By extension, many people consider computational social science the analysis of large computationally-intensive datasets – namely, computational social science centers on those data analyses that require powerful digital machines (e.g., supercomputers) and on the programmers and engineers associated with such machines.

Despite the intuitive appeal, defining computation in terms of digital machines is problematic. Such a definition lacks reasonable stability. It also reduces computation to an instrumental tool rather than a way of thinking (SIPSER 2012): »The computer [as machine] is not the point« (EPSTEIN 2006: xiii). As the power and algorithmic efficiency of computers evolve and as datasets grow larger and more complex, a machine- or data-based definition of computational social science becomes unstable. It requires researchers to keep determining whether, or if, their current project is complex enough for a machine to count as »computational«. This is not to say that big data are irrelevant to understanding computational social science. *Big data* – data of exponential volume, variety, velocity (LANEY 2001), and messy complexity – inspired computational social science because researchers needed new ways to analyze large-scale social data. Yet, big data is not essential to computation. Computation involves more than large-scale »number crunching« of unstructured or multi-structured large datasets. Computation is a way of thinking about research problems – a way of thinking with a long history prior to the development of information technologies as methodological tools and large-scale social datasets (HORSWILL 2012; SIPSER 2012; WING 2006).

Computation is also not just calculation. Defining computation as a way of thinking about research problems moves beyond equating computation with calculation or with the use of numerically-driven algorithms. Yet conflating computation with calculation remains commonplace. Certainly, the predominant use of primarily quantitative approaches is intuitive since computers as machines were invented as »number crunchers« (SIMON 1980: 6264). Digital machines also perform calculations well with mathematical calculations serving as the prototypical algorithm. Plus, prior to the development of computers as machines, the first computers were people (usually women) who performed mathematical calculations. Yet, for many people, computation remains equivalent to mathematical calculation except that machines now perform the »mental operation involving numbers« (HORSWILL 2012: 9). Consequently, there is a long history of connecting computation to calculation.

Although, numbers are often involved in computation and computation has a long history in mathematics, computation need not be constrained to numerical calculation or representation (SIMON 1980; also DENNING 2010). Non-numerical computation is also possible, if less well-developed, than numerical computation: Computers can read, write, compare, copy, and branch instruction and data symbols. Although the symbols analyzed in non-numeric computing are often letters representing the variables of algebraic equations, in theory, complex symbols that are manifest in human language or behavior could also be represented and computed. Such complex symbolic computation challenges the current abilities of existing machines, algorithms, and assumptions (DENNING 2010). But therein lies opportunity. Development of non-numerical computation is growing in computational linguistics, cognitive science, and artificial intelligence, following years of lagging behind the development of numerical computation (HAUSSER 2014). Thus, although computers (as machines) are particularly good at calculation, computation need not be inextricably tied to numbers, nor to the capabilities of the current machines that calculate those numbers. »Computation«, then, »is a kind of question-answering«, a standard of explanation that focuses on the precise and specific processes, conditions, workflows, decisions, and, therefore, algorithms, involved in answering questions about how a phenomenon works (see HORSWILL 2012: 2).

Computation, then, is a particular way of thinking that focuses on how algorithms can be used to represent, model, simulate, identify, analyze, or simulate phenomena in diverse ways. As instances of logic, *algorithms* are self-contained, step-by-step set of precise and rigorously defined operations, rules, and conditions that answer questions and solve problems (SIPSER 2012: 182). Although often developed for digital machines, algorithms existed prior to the advent of modern computing machines. Defining computation as a way of thinking (NATIONAL RESEARCH COUNCIL 2010; WING 2006) is a conceptual rather than technological or numerical move (EPSTEIN 2006). Computation does not necessarily require one to quantify a specific social phenomenon. Rather, it involves applying rules or executing steps in a systematic, specific, logical way to represent, understand, and evaluate different phenomena.

Defining computation as a systematic, algorithmic, particular way of thinking expands understandings of computational techniques as well as social science theories and phenomena. Pragmatically, defining computation as a particular way of thinking expands the types of questions researchers believe computational social science can answer. Such a definition helps people see the potential of computational social science to answer questions in a wide range of numerical and non-numerical problem domains. It avoids the tendency to circumscribe computational understandings of social phenomena to quantitative insights or numerical calculations of big data. Finding computational answers to diverse problem domains might require developing techniques, algorithms, processes, workflows, and hardware that allow a broader range of problem types to be solved using the precise, logical algorithms computation requires. It may also involve determining when and if computation is sufficient, appropriate, or feasible for answering certain research questions given potential datasets; ontological, axiological, and epistemological assumptions; and computational power and processes. Viewing computation as *metatheoretical* – namely, a way of thinking about the world – provides a generative perspective for studying social phenomena. The practical and theoretical requirements of computation differ from the requirements used in conventional social science (EPSTEIN 2006). As a result, researchers need to reconsider research assumptions and conventions. Although often left unspoken, such considerations form the foundation of rigorous and innovative research design, not just computational social science.

In brief, computational social science requires representing complex social phenomena in ways amenable for computing, without reducing computation to numerical calculation. Such representation requirements are needed by the algorithms that underlie computational thinking and analysis. Computational representations often differ from those used in conventional analytic approaches, which were typically conceived and developed with human analysts in mind. Yet, by representing or framing a social phenomenon in different ways, researchers can gain access to conceptual, empirical, and practical insights that any one approach is unlikely to offer. Just because computers (as machines) are particularly good at numerical calculations does not mean that analyses of social phenomena and associated data should be reduced to those questions best answered, or conventionally answered, by numerical calculations. Consequently, computational social science should evolve to consider how the full range of social science methods and their associated entanglements could inspire and inform computational insights and approaches – and by extension the insights that large-scale social datasets can provide.

Computation – as a way of thinking and a method – allows researchers to refine their understanding of the phenomenon in question by requiring them to explicate it clearly as well as by offering access to new types, ranges, and amounts of data that offer evidence of the phenomenon. To accomplish this goal, we encourage starting with curiosity: What is the core question that drives your research? What questions matter to social science? To the publics we serve? Once you have your driving question, only then consider the computational possibilities and requirements needed to answer your question. In the sections that follow, we describe this process. We outline the proposed workflow and illustrate it using an exemplar from one of our current research projects.

2. A Proposed Workflow for Computational Social Science: Using Social Contract Theory as Example

The proposed workflow begins with reflection. In the case of computational social science, this reflective approach ties together the theoretical requirements of the phenomenon and each researcher's metaphysical commitments with the operational requirements of computation. Focusing on reflection as foundational for high-quality research is an old idea. Reflecting on one's metatheoretical assumptions, research questions, expertise, intended audiences, and intended approaches helps scholars determine whether their research team, research questions, intended methodological approaches, and datasets are well-suited to the research problem at hand. Broadly speaking, reflection includes considering strengths, shortcomings, biases, and assumptions of one's research approach. Metatheoretical reflection includes considering what counts as trustworthy evidence of this phenomenon, what counts as real; how, when, and whether one should act on research or advocate for involved parties; and how we come to know about a phenomenon.

Reflection is a hallmark of effective, rigorous social science research regardless of one's methodological commitments, interests, or expertise. Yet, whereas reflection often happens implicitly in quantitative methodological traditions, qualitative methodological traditions tend to celebrate reflexivity as an explicitly recorded practice (TRACY 2010). Memoing, talking, and journaling – individually and collectively – can help identify biases, miscommunications, challenges, and potential opportunities during research design and implementation. A reflexive approach also offers insights into unexpected results, opportunities, and challenges. It helps create an audit trail. Given growing concerns about the lack of reproducibility in social scientific work (OPEN SCIENCE COLLABORATIVE 2015) and the ethical and practical implications of »black box« algorithms, detailed audit trails offer documentation for the myriad decisions that informed the workflow of a research project. Although reflective processes may be invisible in publications, reflection provides a necessary foundation for effective, ethical, generative, and rigorous research (BOYD/CRAWFORD 2012; TRACY 2010).

During the design and development aspects of the research workflow, researchers benefit from reflecting on the theoretical and computational requirements of their project. Emphasizing theoretical requirements focuses attention on the current strengths and weaknesses of theory to explain the mechanisms underlying the target phenomenon. Take time to identify and explicate the implicit and explicit concepts, relationships, and assumptions of relevant theory in as much detail as possible. Explication should also consider each researcher's metatheoretical assumptions and commitments as they are consequential for the practice of research. Detailed explication requires drawing from existing conceptual and empirical research, often from multiple disciplines, and with consideration of multiple methodological approaches that could help answer the question. Explication is more than explanation. Clear explication sharpens connections between theory, data, and method while highlighting implicit assumptions and missing information to generate opportunities for insight (CHAFFEE 1991).

Computation as a way of thinking focuses attention on computational requirements rather than computational constraints. That is, computational thinking defines what computation needs to be able to do, rather than what computation can not do. An initial focus on requirements rather than constraints avoids premature closure based on perceived or actual limits of current computing power or algorithms. It centers research around the social phenomenon in question rather than current machine or algorithmic capabilities. The point is not the machine (EPSTEIN 2006); the point is improved understanding of the social phenomenon: Researchers want to explain, predict, control, intervene, advocate for, or otherwise understand and influence human behavior.

As Figure 1 shows, our workflow starts with curiosity. Here, curiosity takes the form of a motivating question. Initially, this motivating question may be unrefined. The workflow includes three main phases: (1) examining motivations, (2) evaluating evidence, and (3) designing computation. Each phase focuses on considering a set of guiding questions. Reflectively