Computational Communication Science:

Lessons from Working Group Sessions with Experts of an Emerging Research Field

Stephanie Geise & Annie Waldherr

To cite: Geise, S. & Waldherr, A. (2021). Computational communication science: Lessons from working group sessions with experts of an emerging research field. In U. Engel, A. Quan-Haase, S. X. Liu, Su & L. Lyberg (Eds.), *Handbook of computational social science. Volume 1: Theory, case studies and ethics* (pp. 66-82). Routledge.

https://doi.org/10.4324/9781003024583

Abstract

This chapter provides an overview of computational communication science (CCS) as an emerging and exemplary subfield of computational social science. Based on lessons from working group sessions with 34 experts, we address recent challenges and desiderata of CCS research while reflecting upon its future development and expansion. Four major fields of action proved particularly relevant in these discussions: First, challenges related to a reflected but integrated CCS methodology; second, challenges related to a further elaboration on the theoretical perspectives on CCS; third, challenges related to the formation and further institutionalization of CCS as fundamental basis for further scientific exchange, progress and standardization; and fourth, implications for empirical communication research, particularly highlighting the relevance of wicked problems as research incubators driving further progress.

Keywords: computational social science, computational communication science, communication theory, methodology, research ethics, group discussions, wicked problems

Computational Communication Science:

Lessons from Working Group Sessions with Experts of an Emerging Research Field

We live in a world full of digitized communication: Search engines such as *Google*, social networking sites such as *Facebook*, *Twitter* and *Instagram,* mobile applications and digital news portals are integral parts of our lives. The ubiquity of digital communication facilitates our living, keeping us informed and vividly in touch with our peers. At the same time, digital media generates a wealth of sensitive information about ourselves, our social connections and communicative relationships, as well as of our media usage behavior, with which we leave a dense network of individual digital traces. These digital traces open up new opportunities for communication research, particularly to explore social communication phenomena that arise from the use, appropriation and impact of digital technology and media. These phenomena range from questions of private communication in the mediatized home, issues connected to the collection of individual messages and the sharing of health and fitness data, to the critical analysis of the digitized formation of public opinion such as potential filter bubbles, fake news as well as hate speech—to name but a few current examples.

It is therefore only logical that communication and media studies have increasingly focused their knowledge and analytical interest on phenomena of and around digitization. But current developments in communication science are not only promoted by a growing interest in substantive questions related to digitization: Driving forces also include technological advancements in computing capacities, the rise of machine learning, and the ubiquity of artificial intelligence. These trends open up new opportunities for communication researchers to access existing data sets, to collect vast amounts of structured and unstructured data 'on the side', while at the same time new, computational approaches for data analysis are developing, which also facilitate the content-related examination of the data. Accompanied by an

incorporated paradigm shift from traditional empirical research to potentially large-scale computational methods, the research field of computational communication science (CCS) has emerged, responding to the demand to adapt repertoires of analysis and methods to the changing conditions of increasingly digital media and digitized communication (e.g., Hilbert et al., 2019; Niemann-Lenz et al., 2019; Van Atteveldt & Peng, 2018).

Yet, the speed at which computational social science (CCS) is expanding should not hide the fact that it is still a fragmented field in much need of consolidation. Embedded in the broader 'computational turn' in the social sciences, the changes in our field not only offer new opportunities, but also bring new methodological, data-analytical and research-ethical challenges with it. Particularly, the application of advanced, often machine-driven approaches implies serious challenges of this kind (see, for example, Berinsky, Huber & Lenz (2012) on Amazon's Mechanical Turk), which empirical communication research must face.

The ongoing implementation of computational methods has also been criticized for being data-driven and deficient in theoretical positioning, and valid criticism has been made particularly regarding the reliability, validity, and reproducibility of computational methods (e.g., Mahrt & Scharkow, 2013; Waldherr, Geise & Katzenbach, 2019). Especially because of their novelty and their potential to expand the visibility of communicative processes, CCS explicitly needs to be subject to a critical and reflected debate within the social sciences itself.

Aiming to contribute to this discourse, this chapter provides an overview of CCS as an emerging and exemplary subfield of computational social science. Based on lessons from working group sessions with 34 experts, we address recent challenges and desiderates that CCS researchers from communication and political science, informatics, computer science, linguistic, sociology, and other related research fields see while reflecting upon its development and its future expansion. We summarize and further discuss four major fields of action these experts have put particular focus on: First, challenges related to a reflected, but

integrated CCS methodology; second, challenges related to a further elaboration of theoretical perspectives on CCS, and third, challenges related to the formation and further institutionalization of CCS as fundamental basis for further discourse, quality control, and scientific progress; and fourth, further ideas as to how the challenges discussed can be tackled and implemented by empirical communication research and which concrete decisions should prove to be relevant in this context.

The Computational Turn of Communication Science and its Characteristics

At the heart of the digitalization of society is digitized communication. Communication scholars have been quick to point out that major parts of the digital traces available in digitized information environments are factually communication data (Shah, Cappella & Neuman, 2015)—such as text messages in chats and social media posts, memes, podcasts or audio-visuals—or data on media use and exposure— such as tracking, log, or search data. Consequently, with the key role communication plays in the increasing digitalization of contemporary societies, computational communication science (CCS) has emerged as one of the most vibrant subfields of computational social science (CSS). It therefore makes sense to take a closer look at CCS as an exemplary—and at the same time increasingly important—research field of CSS.

With Hilbert et al. (2019, p. 3914), we understand CCS "as an application of computational science to questions of human and social communication." Although the rapid development of the field has certainly been triggered and catalyzed by the massive availability of digital traces (Hilbert et al., 2019; Shah et al., 2015), big data is only one of the defining instances of CCS. Another defining characteristic of CCS is its strong orientation towards computational methods to investigate social communication phenomena. This has been driven by quickly growing computing capacities enabling powerful methods for

modeling large and complex data. Over the last decade, communication scholars have increasingly taken advantage of computational approaches such as automated content analysis, network analysis, or computer simulation to answer fundamental questions about human behavior, interaction and communication. From a topical perspective, CCS researchers focus primarily, but not exclusively, on interactive computer-based phenomena and the associated technologies, structures and processes of computer-mediated reception, information processing and communication as well as their social consequences.

The powerful combination of big data and computational methods has led scholars to predict an "unprecedented boost to progress" for communication science (van Atteveldt & Peng, 2018, p. 82). In recent years, this development has gained momentum. Several programmatic articles on the potentials of computational methods in communication research have been published (Choi, 2020; Hilbert et al., 2019; Niemann-Lenz et al., 2019; Van Atteveldt & Peng, 2018). In 2016, a new interest group on computational methods in the International Communication Association (ICA) was founded, which only four years later would be established as a proper division of the association. In 2019, the first issue of a new open-access journal, Computational Communication Research*,* was published.

Apart from these first instances of institutionalization on an international level, CCS is still in its formative stage, and a lot of open questions merit a thorough discussion. For example, while the potential of computational methods for data analysis is obvious, the role of theory in this emerging field is still rather unclear and underrepresented (Waldherr, Geise, Mahrt, Katzenbach & Nürnbergk, 2021). It seems that the rapid development of computational methods has not been accompanied by an equally strong emphasis on theoretical developments within the scholarly community. Which theories and theoretical concepts are available that can enrich the methodological discussion on CCS? How can

theory development inspire further methodological advances? What are the main challenges and core issues to be addressed in the further development of CCS?

Working Group Sessions with Experts: Line-Up and Learnings

All of these questions were taken up and openly discussed in the context of an interdisciplinary working group "Computational Communication Science 'en route': theory, methodology & research ethics" founded by the German Center for Applied Internet Studies (CAIS). The 34 invited experts in this working group came from the research fields of communication and political science, computer science, informatics, linguistics, sociology and other related research areas. During the funding period of around one year, we met regularly in collaborative workshops, reflecting on the development of CCS, intending to assess its future advancement and potential ways of improvement. Guided by our key questions and conceptual impulses, the discussions revolved around the interplay between subject-related interests and their underlying driving research questions, and the partly highly specific theoretical and methodological perspectives of CCS, its possible interfaces with established social theories and the methodological questions and advances that arise from them.

Methodologically, the group sessions were conceptualized as a combination of explorative expert interviews and dialogue-oriented, transdisciplinary workshops (Bergmann, Jahn, Knoblauch, Krohn, Pohl & Schramm, 2010; Defila & Di Giulio, 2014; Dexter, 1970; Niederberger, 2014). The aim was to encourage the invited experts to engage in an open, collaborative exchange in order to inquire about their expert knowledge, assessments, experiences, ideas and development perspectives, and to involve them in a further, constructive dialogue (Defila & Di Giulio, 2014; Dexter, 1970) about the future of CSS/CCS. At the same time, the invited experts were themselves part of the social field of action of CSS/CCS we aimed to explore.

From our perspective, there were three arguments in favor of establishing a workshop series that was situated in the field of communication science, but was equally open to neighboring disciplines in the social sciences: Firstly, collaborative expert workshops are particularly well suited for exploring and eliciting lines of development, potential problem areas as well as the resulting requirements and solution approaches, because they stimulate an interactive and creative exchange of multi-faceted expert experience (Defila & Di Giulio, 2014; Niederberger, 2014). Second, although straight methodological guidelines to systematize the elicitation process of experts are rare (Defila & Di Giulio, 2014; McDonald, Bammer & Dean, 2009), the involvement of experts in social science research is becoming increasingly important (Niederberger, 2014). In our case, the principal motivations for our workshop approaches were to reduce the complexity of the research process, improve the quality of the results, and address some of the major issues often experienced by computational researchers. Third, at the national level, or rather in the German-speaking countries, researchers working with computational methods often are regarded as 'lone wolves', because institutional opportunities for interdisciplinary networking are still lacking. With the establishment of the working group as well as with this article, we thus not only aim to gain in-depth knowledge, but to extend the expert discussions to the larger CSS/CCS community, finally intending to contribute to its further institutionalization.

Having these lofty goals, we conducted a series of guided collaborative CCS/CSS-expert workshops, thematically ranging from 1) challenges related to a reflected, but integrated CCS *methodology*, 2) challenges related to a further elaboration of specific *theoretical perspectives* on CCS, and potential interfaces to established social science theories, and 3) challenges related to the *formation and further institutionalization* of CCS as fundamental basis for further discourse, quality control, and scientific progress. Focusing the discussions along the lines of an exemplary thematic field of research— namely the

digitization of public opinion formation— we 4) further discussed ideas as to how the identified challenges could be tackled and implemented by empirical communication research, and which concrete decisions should prove to be relevant in this context. In the following, we give an overview of the results of these experts' workshops and integrate them into the existing CCS research landscape.

Learning 1: Contours of an expanding field of research

In our kick-off workshop we aimed at developing a common understanding of the research area, orientation and objectives of CSS and CCS. The experts agreed that CSS is not yet a social-scientific discipline on its own, but may be characterized as a "fluid network" or "movement" within the social sciences. Some experts also suggested that CSS—in its specific orientation—could serve as "auxiliary science" for the social sciences, just like statistics; and some even noted that CSS is a "buzzword" or "label" that might be used strategically to position researchers and institutions in competition for resources, sometimes even without having any substantial meaning.

In the discussion, we did not follow this last thought, but identified three components we deemed essential for defining the field of CSS: (1) the analysis of social processes (2) with (big) digital data on these processes (3), and the use of computer-based methods that allow a high degree of automation of the research process. Later, we specified these dimensions as "driving forces" of CCS (Waldherr et al., 2021). Just as in the social sciences in general, the big (communication) data deluge and powerful computing capacities have been main catalyzers for the computerization and algorithmization of communication research (Hilbert et al., 2019). Although digital data and computational methods can be applied to any research question in communication, a third driver of CCS—we and others have identified—are often highly complex research problems of human communication which newly emerge in digitized

societies, and which call for developments in theories as well as methodological toolkits (Niemann-Lenz et al., 2019), thinking, for example, of phenomena such as filter bubbles, Twitter storms, flash mobs, cyber-bullying, etc. (so-called "wicked problems"; see subsequently).

CSS in general is a highly interdisciplinary field. Yet, the impression of our experts was that CSS is currently dominated by the technical disciplines such as computer science, engineering, mathematics, and physics. Consequently, the field is experienced as highly method- and data-driven. For example, a number of interdisciplinary conferences have a pronounced bias towards computer science (e.g., Social Informatics, Complex Networks, ICWSM, IC2S2) or a strong focus on one specific method (NetSci, Sunbelt, ESSA). Social scientists and their theoretical approaches as well as methodological standards thus do not appear to be represented equally and "have come late to the party", as one of our experts phrased it.

This is partly also true for the specific field of CCS. However, communication researchers quite early opened up to computational approaches themselves, engaged in further developing computational methods for their purposes and started scholarly discussions in forums quite central to their discipline (e.g., in the computational methods division of the ICA, and in several special journal issues addressing the field; Domahidi, Yang, Niemann-Lenz & Reinecke, 2019; Theocharis & Jungherr, 2021; Van Atteveldt & Peng, 2018). Thus, the computational turn in communication science by now appears to be well embedded into the overall discipline, although "far from being normalized" or "mainstreamed". In our perspective, CCS can really profit by having trained communication researchers who are skilled in computational approaches, but also savvy in communication theories. This potentially is an advantage compared with other social-scientific disciplines, as for example

political science where scholars have observed that internet research and data science is still regarded as a "playground of cute nerds" (Jungherr, 2017, p. 301).

Learning 2: Theories in CCS

In the context of the euphoria in the early development stage of CCS, some colleagues have argued that the availability of large data sets and computational methods would make theories obsolete (Anderson, 2008; Mayer-Schönberger & Cukier, 2013). This positivist view was widely criticized (Driscoll & Walker, 2014; Mahrt, 2018; Mahrt & Scharkow, 2013; Vis, 2013)— and we as well took it as a pointed starting point for a further discussion of the role of theory in CCS for which we devoted the second, two-day expert workshop. Starting from an exploration of potentials of a theoretical "common ground" of CSS and its major challenges, the aim of the theory workshop was to elaborate existing or imaginable theoretical references and interfaces of CCS research as well as to identify requirements of appropriate theories that can inspire and advance it.

Our analytical interest was guided by three main considerations we initially carved out in the workshop: First, as other scholars before, the invited experts shared the belief that even exploratory pattern detection is necessarily driven by numerous theoretical assumptions that are reflected in how data are collected, analyzed, and interpreted (Andrejevic, 2014; Crawford, 2013). Second, and interconnected, the invited experts agreed that research that lacks sufficient theoretical reflection of the studied phenomena risks producing methodological and/or data-analytical artifacts instead of meaningful findings about social reality (Bright, 2017; Kitchin, 2014; Ruths & Pfeffer, 2014). CCS scientists have thus, and third, demonstrated that computational methods in many respects are ideally suited to empirically test both long-term social theories (González-Bailón, 2017) as well as medium-term communication theories such as agenda-setting (e.g., Vargo, Guo, & Amazeen, 2018);

and computational methods have also been used to inductively develop and advance theories in a computational grounded theory approach (Choi, 2020).

However, in our view a discussion about CCS and theory, a larger reflection of how theories, particularly meta-theoretical frameworks, can serve CCS scholarship, is widely missing. But how can social science theories advance CSS/CCS? And which theories are particularly suitable for empirical theory-based CSS/CCS research? To discuss these questions, in the second workshop stage, we conducted a "CCS theory roadshow", for which each of the invited CCS experts had prepared a theory input, which was then presented and discussed in a round table format. For this input, the invited experts a priori selected theories that they experienced highly suited for research in the field of CSS and prepared "theory profiles" guided by the following evaluation scheme: 1) short description of the theory/theoretical framework (including basic ideas, core assumptions, central concepts and premises); 2) disciplinary setting and previous application (e.g., in which fields of research, for the analysis of which phenomena, in the context of which research contexts); 3) gain in knowledge and/or analytical potential for CSS/CCS (e.g., specific theoretical perspectives, new questions, sensitization for otherwise unilluminated structures, processes or phenomena); 4) challenges and ambiguities (e.g., premises theoretically worthy of discussion, empirical verifiability of central assumptions, contradictions); 5) possible interrelations and/or conflicts with other established theories relevant to CSS/CCS.

On that basis, we worked out inspiring theoretical grounds and interconnections of the different approaches, identified explicit and implicit premises of CCS research guided by theory, and derived general requirements that theories should meet in order to provide further insights for CCS research (see Table 1). Along these criteria, we discussed the potential of five theoretical frameworks in greater depth and intensity, namely: 1) structuration theory, 2) actor-network theory, 3) complexity theory, 4) mediatization and 5) theories of the public

sphere. The discussion carved out that the selected theories can indeed constructively inform and inspire CCS research—although the different concepts naturally involve different emphases, challenges and limitations.

Table 1: Premises of CCS research and requirements for theories in CSS

Premises of CCS research	**Requirements for theories**
Conclusions about social reality can be drawn from digital traces.	Consideration of **digitization and media change**
Media change and digital technologies influence social reality.	Analysis of the **interdependencies between social structures, individuals and technology**
Networks describe and explain social structures.	Connection of several levels of analysis
Human-machine interactions are a central subject of CSS/CCS research.	Options to model **dynamic processes**
Findings lay claim to truth, objectivity and completeness.	Raising **awareness of hidden bias** in the data and normativity of research
CSS/CCS helps to find technical solutions for social problems.	**Application for empirical analysis** with computational methods

We then summarized our findings and reflections at a higher level. Here it became clear that CCS experts saw particular benefit in applying macro frameworks to the analysis of CCS/CSS phenomena. Even though macro-theoretical frameworks may not directly point computational scientists to any particular research questions or hypotheses, they do have a lasting impact on their research: theoretical frameworks strongly influence what types of questions we ask, what types of hypotheses we formulate, and what (computational or traditional) methods we consider appropriate and useful to investigate them. A conscious foundation of our thinking and of (computer-aided) communication research through macro-theoretical frameworks forces us to make explicit and reflect the epistemological basic assumptions of our work (Resnyansky, 2019). In addition, macro-theoretical frameworks inspire us to answer critical questions, such as: How do we as CCS researchers understand the world? How do we try to

decipher meaning and significance in this world? What is our research interest (explanation, description, standardization or criticism)? Which actors and/or structures do we consider relevant to explain social communication processes based on the computerized collection and/or analysis of (digital) data?

The reproduction of the complete theory discussion would go beyond the scope of this book chapter. However, our considerations are documented elsewhere: In a first theory paper, we devoted ourselves more intensively to the discussion of the potentials and possible applications of the actor-network theory (ANT), revisiting the framework as a helpful concept for theorizing and inspiring CCS scholars' research agendas (Waldherr, Geise & Katzenbach, 2019). Based on a literature review, we showed that ANT has already been used to study a variety of questions in (computational) media and communication research, and identified three key contributions of an ANT-perspective in CCS: First, by focusing on the role of technology in communication and the relationships between nonhumans and humans guided by ANT, the framework helps us to open our discipline to new questions and perspectives of communication research. Second, from a meta-perspective, the framework suggests to further reflect on the emergence, development, and inscriptions of computational methods in CCS, as they are not only shaped by technological and scientific innovations, but they also influence how we do communication science now and in the future. And third, this meta-perspective motivates an even further step of self-reflection, pushing researchers to deliberate about their roles in the research process as well as the normative and ethical assumptions guiding them. Taken together, as we illustrate in the article in more detail, the outlined theoretical considerations open up potential future perspectives of ANT-informed CCS.

In addition to this contribution devoted to a specific theoretical conception, we dedicated a second article to the added value of *macro-theoretical perspectives*— such as, for example, *complexity theory*, *theories of the public sphere*, and *mediatization theory*—in

establishing a meta-theoretical underpinning of CCS research (Waldherr et al., 2021). With the help of three analytical dimensions—1) interdependencies between varieties of entities and actors on different levels of hierarchy and organization, 2) the consideration of normativity and 3) multi-level dynamics in the research process—we argue that CCS scholars can benefit from connecting their empirical, often highly innovative work to established macro-theoretical frameworks: Particularly because these frameworks make explicit how computational research foci (and blind spots) and designs are shaped by (implicit or explicit) theoretical underpinnings, and how these significantly impact computational research questions, hypotheses, and methods, the integration of macro-theoretical perspectives can inspire and advance future CCS scholarship.

Learning 3: Methods of CCS

In the third expert workshop, we took two days for in-depth discussions of several selected computational methods. When choosing the methods of interest, we aimed for a broad variety of approaches. Besides methods of digital trace data analysis, which are already well established in communication research— such as network analysis, text classification, web scraping, web tracking, and bot detection— we also discussed methods that have not been in the center of the development so far, but in our view add important perspectives to CCS, namely agent-based modeling as well as computational observation methods such as eye tracking and automated facial emotion recognition. Of course, this set of methods also is by no means exhaustive and represents only a small selection of the current toolbox of CCS. Yet, all of these methods are computational in that sense that they are able to analyze and/or generate large amounts of digital data and enable a high degree of automation of the research pipeline (Jünger & Schade, 2018). In addition, they belong to the realm of communication science as they are used to study processes of human communication.

On the basis of short methodological profiles, the CCS/CSS experts had prepared in advance, we discussed the specific insights gained by each method as well as the methodological and ethical challenges of the methods, which were perceived as central in the current debates by the experts. The essence of these discussions is summarized in Table 2.

Independent of the methods discussed in each case, nearly all of our experts reported challenges concerning the *representativeness* of CCS data. At the heart of this challenge is the question of how well the data or the model "represent the real social world", and how safe it is to draw conclusions based on this data or model. Issues under discussion ranged from questions of boundary definition and drawing representative samples from social networks to technical and social biases in data.

Technical biases might occur, for example, due to access restrictions (through APIs and their terms of services, or bot blockers on web pages), or technical disruptions in data collection (e.g., through interrupted internet connections or other technical issues). Another form of technical bias is the automated production of social media content through bots. Depending on the research question, detecting and filtering out bot-produced content is essential to prevent biased conclusions. Social biases arise particularly for all research relying on social media data, as these platforms are known to under-represent certain demographic groups (e.g., gender and educational biases on Twitter and Facebook; Barberá & Rivero, 2015; Mellon & Prosser, 2017). Another challenge to representativeness occurs with very demanding forms of data collection such as web tracking or eye tracking which for different reasons often result in very small samples, but with large amounts of fine-grained individual-level data (Möller, Van de Velde, Merten & Puschmann, 2020; Geise, Heck & Panke, 2020).

Table 2: Challenges of CCS methods and desiderata for the CCS community

Challenges of CCS methods	Desiderata for CCS community
Representativeness: drawing samples, access to data, technical & social biases	Ensure access to data, establish data infrastructure
Validity: substantive meaning of measures, defining gold standards & ground truths	Systematic and comparative methods research, definition of best practices
Hypothesis testing: Violation of distributional assumptions, meaningfulness of significance tests with large-scale data	Develop standards for hypothesis testing with large-scale online data
Ethics & privacy: sensitive & private user data, difficulties of secure anonymization	Define common ethical standards & techniques for secure data sharing

Another common theme in the expert discussion was the *validity* of CCS methods. The core question with regard to validity is matching the measurements with the theoretical constructs under study. While computational methods often produce massive amounts of data, it is not always clear what these data really mean substantively, and what conclusions from it might be drawn or not drawn. For example, there are multiple algorithms for community detection, which are able to identify subgroups in large-scale networks. As Stoltenberg, Maier and Waldherr (2019) have shown, each of these algorithms generate possibly different results in a way that it is important to make theoretically informed choices on which theoretical types of communities in the theoretical sense researchers want to measure (e.g., based on similarity, ideological association, or strategic alliance of actors). In automated text analysis and classification tasks, pressing questions of validity are developing the right gold standards and ground truths for training algorithms (Song et al., 2020), especially when dealing with ambiguous categories.

Comparable challenges arise in the computational observation of sensory behavioral data, as is typically applied in eye tracking and computer-based biometric emotion

measurement. Here, too, the measurement and automated classification of the data is determined by the algorithms used to detect the measured constructs. In eye tracking, for example, a so-called fixation filter is set, so that the processing script is technically able to identify eye-movements that should be grouped together as foveal fixations or saccades. While such a fixation filter is a common and necessary technological requisite, the fixation-classification algorithms used can affect data precision (Hessels, Kemner, van den Boomen & Hooge, 2016). It is also common practice here to subsequently define so-called "areas of interest"—specific areas of the received stimuli, over which the eye tracking data are then compared (Geise, Heck & Panke, 2020; Hessels et al., 2016). Such methodological decisions should critically reflect the fact that standards for setting this technological pre-structuring are only slowly established—but the resulting data and findings are rarely questioned later (Geise, 2014).

The question of ground truth becomes particularly virulent, when it is socially and politically consequential, such as in bot identification. While detecting and filtering out bots from social media datasets might increase the representativeness and validity of the results, bot detection algorithms such as botometer are prone to produce false positives (Keller & Klinger, 2019). Because bots constantly evolve and are a "moving target", it is also hard to define a reliable ground truth against which to test the algorithms. Finally, in computational modeling the validity of basic assumptions and parameter choices as well as finding an appropriate level of abstraction is key (Sun et al., 2016). Particularly, strategies of empirical validation are an issue of ongoing scholarly debate (Gräbner, 2018).

Digital trace data generally are a lot more messy than the data gathered in classical empirical research (Waldherr, Maier, Miltner & Günther, 2017). For example, online data, and particularly network data tend to produce highly skewed distributions and nonlinear relationships (Adamic & Huberman, 2000; Broido & Clauset, 2019), which prohibits the

unreflected use of statistical modeling based on normal distributions and linearity assumptions (such as ordinary least square regressions). This makes *hypothesis testing* complicated, requiring rather complex techniques of analysis (such as *Exponential Random Graph Modeling* in network analysis; Lusher, Koskinen & Robins, 2013). Because this considerably raises the analytical bar for communication scholars (and reviewers), this often leads to rather descriptive and exploratory studies. Additionally, the mere scale of data produced in CCS studies often makes the use of inferential statistics and significance tests obsolete--an issue which has also been discussed for the large amounts of data generated in computer simulations (Troitzsch, 2016).

The fourth and final set of challenges, the invited experts highlighted centered on questions of *ethics*, and particularly *privacy*. When working with fine-grained individual-level data, such as in web tracking studies, highly sensitive user data is generated. This requires comprehensive privacy protection protocols ensuring not only informed consent before installing tracking apps, but also that only necessary data is collected (e.g., by defining white and black lists of URLs for data collection) and that data is securely anonymized (Möller et al., 2020). Of course, meeting these standards is impossible if researchers work on problems such as bot detection and misinformation where it is essential to collect data without users' consent, and also access to data deleted by users would be helpful.

Based on the discussions of these challenges, we derived desiderata for the development of CCS methods and the CCS community. It was suggested to create systematic overviews of research questions and adequate computational methods and software packages (e.g., in the form of a handbook or an electronic tool collection). The development of best practice standards and more methodological research in the form of systematic comparisons of methods and instruments was also highlighted as necessary. For the community, a regular and institutionalized interdisciplinary exchange was proposed; and problems of equal data

access, establishing data infrastructures and protocols for data and code sharing were addressed.

In sum, it is reassuring that the major CCS challenges largely correspond to those of classical empirical research. However, they come in new qualities so that long established solutions cannot simply be adopted. This implies a high potential of making CCS research frustrating because researchers first need to solve very basic methodological problems before they can turn to answering their primary research questions. However, the expert discussions made clear that one of the biggest CCS desiderata is precisely this: advance systematic, comparative methodological research, and develop common standards for data collection and data analysis.

"CCS en route": Central Challenges and Future Perspectives for an Emerging Research Field

The last workshop aimed at an integrative synthesis of the previous findings, on the basis of which we wanted to work out the question of future perspectives and challenges of CCS research. We proceeded in two steps: In step 1, we summarized our findings from the previous three workshops in a few bullet points, condensed the central arguments and, together with the invited experts, specified which desiderata arise with regard to method and theory development in CCS. On this basis, in step 2, we worked out which next steps should necessarily result from our analysis of the current status quo and how and where we, as CCS researchers, can contribute to the implementation, further professionalization and institutionalization of CCS research and scholarship. Our central learnings can be condensed into four conceptual statements the experts agreed with, which we summarize below:

The virtue of disciplinary openness and interdisciplinary collaboration

From the very beginning, our collaborative expert workshops were based on disciplinary and thematic openness. Leading to multi-faceted, inspiring and highly interesting discussions the workshop series again highlighted that we, as communication scholars, can immensely profit from disciplinary openness and interdisciplinary collaboration, particularly with, for example, political science, sociology, computer science, and linguistics. All experts emphasized that an interdisciplinary approach in small workshop groups proved to be a particularly effective format and should be maintained in subsequent projects.

The benefit of sharpening the CCS focus

At the same time, it became apparent that we achieved particular analytical depth when the questions and discussions became very specific and case-related. For this reason, the invited experts advocated a content-based focus and inner-disciplinary bundling of the discussion—this can be interpreted as a commitment to the further establishment and institutionalization of a specified CCS, whereby all experts expressed the wish for a "home" under the umbrella of the larger CSS. Keeping a focus on specific topics, problems and examples seemed to ease further progress in the discussion and work on specific solutions.

To give an example, the opportunities and challenges of CCS were unfolded in a prototypical way with the example of forms, functions and processes of digitized opinion formation, which is becoming increasingly relevant in times of ongoing digital media change. This leads us to the conviction that computational research in communication science is particularly promising where researchers share their research questions, their analytical framework as well as their examples of application. For future work, we therefore suggest a focus on more concrete research problems (particularly on 'wicked problems', see below) in order to discuss and develop specific theoretical and methodological solutions for concrete theoretical and methodological problems at hand.

The potential of a stronger theoretical grounding of CCS

One of the recurring questions the experts raised in the workshops was how we—as social scientists—can add value to the further development of the research field. What contribution can communication scientists make to the debate of an expanding computational research at large? What can the field really profit from in its further development? In view of the discussions about the lack of theory in CSS/CCS research and the frequently observed strong application orientation in the field, the strengthening of the theoretical foundation crystallized as an important developmental step to which CCS experts in particular can contribute meaningfully.

The need for institutionalization

All participants agreed that a focused, interdisciplinary collaboration is easier, more interactive and more productive in small groups, and expressed the wish for a continuation and perpetuation of the exchange that had begun with our workshop series. At the same time, it became apparent that the institutionalization of CCS research is experienced as both, potentially one of the biggest drivers as well as one of the biggest obstacles to further development—especially if the hitherto low level of institutionalization is not further expanded and promoted. In order to create an opportunity for cross-locational, topic- and task-related exchange and thus promote further development of CCS research, the establishment of CCS scholars' networks seems an effective, next step to proceed. The importance of stronger networking for the exchange of experience was substantiated by the experts, among other things, by the example of shared tools, methods and procedures of analysis, but also with regard to an overarching "best practice" platform to facilitate

interdisciplinary cooperation, and to increase the reproducibility, replicability, and generalizability of findings.

In these considerations it is already indicated that the invited experts strongly appreciated the development towards an open science—considerations that have been the subject of intensive science policy discussions, particularly in recent times (e.g., Dienlin et al., 2020; Niemann-Lenz et al., 2019; Nosek et al., 2015). At the same time, the experts were very much aware that many content- and data-sharing activities connected with "best practice" exchange also are associated with additional ethical and legal challenges - including, for example, the acknowledgement of privacy requirements and terms of services, as well as an increased risk of data misuse by third parties. As in similar contexts (Niemann-Lenz et al., 2019), the experts pointed out that the additional effort required to comply with the necessary ethical and legal standards of open CCS research (e.g., complete anonymization of very large amounts of data) could, in the worst case, outweigh the (individual) benefits (e.g. reproducibility). There was agreement, however, that the community should create appropriate framework conditions that promote an overarching "open science" exchange, as this would ultimately benefit the entire development of the research field.

Wicked Problems as Incubators for CCS Research

To conclude, one of the key insights we gained from our workshop series is that "wicked problems" serve as inspiring incubators for CCS (and CSS alike). We borrow the term from public policy research (Rittel & Webber, 1973; Weber & Khademian, 2008) to signify complex and unstructured problems entailing myriad and highly interdependent subproblems which are impossible to be solved without generating multiple threads of follow-up problems. It is exactly this type of problem that many CCS scholars set out to study: big, highly

complex and wicked social problems where communication technology might be the cause and/or solution.

For example, many phenomena which might be subsumed under the term of digitized opinion formation and which have gained much scholarly attention in recent years (Bennett & Pfetsch, 2018; Quandt, 2018) are certainly wicked: e.g., misinformation, fragmentation, propaganda, conflict, or radicalization. Tackling these challenges is a strong motivation for researchers to build interdisciplinary collaborations and overcome methodological obstacles, and it is in these contexts where we observe and expect the most cutting-edge advancements in CCS. Thus, our most important lesson is that we should use this momentum and further connect the different research groups working on similar substantial problems to learn from each other, join efforts on methods development, and eventually help solve the big issues of our digitized times.

Acknowledgements

We would like to thank the CAIS - Center for Advanced Internet Studies (Bochum, Germany) for the generous support, the competent guidance through our funding period and the great hospitality that we and all invited experts enjoyed at the CAIS in Bochum. Without the personnel, organizational and financial support of the CAIS we would not have been able to realize the described workshop series this review is based on. We would also like to thank all the experts -- listed in the appendix -- who discussed the diverse development perspectives of CCS with us intensively and with great passion. Our project would not have been possible without your constructive, cooperative and always inspiring exchange: Many thanks to you!

Appendix: List of Participants

1. Gioele Barabucci, Fellow at the CAIS
2. Matthias Begenat, CAIS
3. Ralf Benzmüller, G DATA SecurityLabs
4. Laura Burbach, RWTH Aachen University
5. Anja Dieckmann, GfK Nuremberg
6. Emese Domahidi, TU Ilmenau
7. Frederik Elwert, Ruhr University Bochum
8. Uwe Engel, University of Bremen
9. Christoph Engemann, Fellow at the CAIS
10. Lena Frischlich, WWU Münster
11. Volker Gehrau, WWU Münster
12. Stephanie Geise, WWU Münster, Workshop Organizer
13. Andreas Jungherr, University of Constance
14. Christian Katzenbach, Alexander von Humboldt Institute for Internet and Society Berlin
15. Melih Kirlidog, Fellow at CAIS
16. Ulrike Klinger, Freie Universität Berlin
17. Iris Lorscheid, TU Hamburg-Harburg
18. Seraphine Maerz, University of Freiburg
19. Merja Mahrt, HHU Düsseldorf
20. Dennis Michels, University Duisburg-Essen
21. Judith Möller, University of Amsterdam
22. Linda Monsees, Fellow at CAIS

23. Christoph Neuberger, LMU Munich
24. Christian Nuernbergk, LMU Munich
25. Cornelius Puschmann, Hans Bredow Institute Hamburg/ University of Bremen
26. Hermann Rotermund, Fellow at the CAIS
27. Andreas Scheu, WWU Münster
28. Armin Scholl, WWU Münster
29. Geeske Scholz, University of Osnabrück
30. Tatjana Scheffler, University of Potsdam
31. Annie Waldherr, WWU Münster, Workshop Organizer
32. Katrin Weller, GESIS Cologne
33. Martin Welker, HMKW University for Media, Communication and Economy, Cologne
34. Gregor Wiedemann, University of Hamburg

References

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science*, *287*(5461), 2115. https://doi.org/10.1126/science.287.5461.2115a

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

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*(1), 1673–1689. Retrieved from https://ijoc.org/index.php/ijoc/article/view/2161

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, *33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Bennet, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of Communication*, *68*(2), 243–253. https://doi.org/10.1093/joc/jqx017

Bergmann, M., Jahn, T., Knoblauch, T., Krohn, W., Pohl, C., & Schramm, E. (2010). *Methoden transdisziplinärer Forschung. Ein Überblick mit Anwendungsbeispielen*. Campus.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Bright, J. (2017). "Big social science": Doing big data in the social sciences. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage handbook of online research methods* (pp. 125–139). Sage.

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*(1), 1017. https://doi.org/10.1038/s41467-019-08746-5

Choi, S. (2020). When digital trace data meet traditional communication theory: Theoretical/methodological directions. *Social Science Computer Review*, *38*(1), 91–107. https://doi.org/10.1177/0894439318788618

Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review Blog*. Retrieved from www.hbrianholland.com/s/05-The-Hidden-Biases-in-Big-Data-Crawford.pdf

Defila, R., & Di Giulio, A. (2014). Methodische Gestaltung transdisziplinärer Workshops. In M. Niederberger, & S. Wassermann (Eds.), *Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschun*g (pp. 69–93). Springer VS.

Dexter, L. A. (1970). Elite and specialized interviewing. In J. A. Robinson (Ed.), *Handbooks for research in political behavior*. Northwestern University Press.

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., . . . Huskey, R. (2020). An agenda for open science in communication. *Journal of Communication.* https://doi.org/10.1093/joc/jqz052

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Outlining the way ahead in computational communication science: An introduction to the IJoC special section on "computational methods for communication science: Toward a strategic roadmap". *International Journal of Communication*, *13*, 3876–3884.

Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication*, *8*(1), 1745–1764. Retrieved from https://ijoc.org/index.php/ijoc/article/view/2171

Geise, S. (2014). Eyetracking in media studies. Theory, method and its exemplary application in analyzing shock-inducing advertisements. In F. Darling-Wolf (Ed.), *The*

international encyclopedia of media studies. Research methods in media studies (pp. 419–444). Blackwell.

Geise, S., Heck, A., & Panke, D. (2020). The effects of digital media images on political participation online: Results of an eye-tracking experiment integrating individual perceptions of "photo news factors". *Policy & Internet*. https://doi.org/10.1002/poi3.235

González-Bailón, S. (2017). *Decoding the social world: Data science and the unintended consequences of communication.* MIT Press.

Gräbner, C. (2018). How to relate models to reality? An epistemological framework for the validation and verification of computational models. *Journal of Artificial Societies and Social Simulation*, *21*(3), 8. https://doi.org/10.18564/jasss.3772

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016). The area-of-interest problem in eye tracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods, 48*(4), 1694–1712. https://doi.org/10.3758/s13428-015-0676-y

Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., . . . Zhu, J. J. H. (2019). Computational communication science: A methodological catalyzer for a maturing discipline. *International Journal of Communication*, *13*, 3912–3934.

Jünger, J., & Schade, H. (2018). Liegt die Zukunft der Kommunikationswissenschaft in der Vergangenheit? Ein Plädoyer für Kontinuität statt Veränderung bei der Analyse von Digitalisierung. *Publizistik*, *63*(4), 497–512. https://doi.org/10.1007/s11616-018-0457-6

Jungherr, A. (2017). The Internet in political communication: State of the field and research perspectives. *Politische Vierteljahresschrift, 58*(2), 284–315. https://doi.org/10.5771/0032-3470-2017-2-284

Keller, T. R., & Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, *36*(1), 171–189. https://doi.org/10.1080/10584609.2018.1526238

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1). https://doi.org/10.1177/2053951714528481

Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications.* Cambridge University Press.

Mahrt, M. (2018). Big data. In P. M. Napoli (Ed.), *Mediated communication* (pp. 627–642). Mouton De Gruyter.

Mahrt, M., & Scharkow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, *57*(1), 20–33.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work, and think*. Murray.

McDonald, D., Bammer, G., & Dean, P. (2009). *Dialogue tools for research integration.* ANU E Press, The Australian National University.

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008. https://doi.org/10.1177/2053168017720008

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*38(5), 616–632. https://doi.org/10.1177/0894439319828012

Niederberger, M. (2014). Methoden der Experteneinbindung. In M. Niederberger & S. Wassermann (Eds.), *Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung* (pp. 33–47). Springer VS.

Niemann-Lenz, J., Bruns, S., Hefner, D., Knop-Huelss, K., Possler, D., Reich, S., . . . Klimmt, C. (2019). Crafting a strategic roadmap for computational methods in communication science: Learnings from the CCS 2018 conference in Hanover. *International Journal of Communication*, *13*, 3885–3893.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., & Yarkoni, T. (2015). Promoting an open research culture. Science, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Quandt, T. (2018). Dark participation. *Media and Communication*, *6*(4), 36–48.

Resnyansky, L. (2019). Conceptual frameworks for social and cultural big data analytics: Answering the epistemological challenge. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951718823815

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063–1064. https://doi.org/10.1126/science.346.6213.1063

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 6–13. https://doi.org/10.1177/0002716215572084

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., . . . Boomgaarden, H. G. (2020). In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication*, online first. https://doi.org/10.1080/10584609.2020.1723752

Stoltenberg, D., Maier, D., & Waldherr, A. (2019). Community detection in civil society online networks: Theoretical guide and empirical assessment. *Social Networks*, *59*, 120–133. https://doi.org/10.1016/j.socnet.2019.07.001

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., . . . Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. *Environmental Modelling & Software*, *86*, 56–67. https://doi.org/10.1016/j.envsoft.2016.09.006

Theocharis, Y., & Jungherr, A. (2021). Computational social science and the study of political communication. *Political Communication, 38*(1-2), 1-22. https://doi.org/10.1080/10584609.2020.1833121

Troitzsch, K. G. (2016). The meaningfulness of statistical significance tests in the analysis of simulation results. *International Journal of Agent Technologies and Systems*, *8*(1), 18–45. https://doi.org/10.4018/IJATS.2016010102

Van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92. https://doi.org/10.1080/19312458.2018.1458084

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, *20*(5), 2028–2049. https://doi.org/10.1177/1461444817712086

Vis, F. (2013). A critical reflection on big data: Considering APIs, researchers and tools as data makers. *First Monday*, *18*(10). https://doi.org/10.5210/fm.v18i10.4878

Waldherr, A., Geise, S., & Katzenbach, C. (2019). Because technology matters: Theorizing interdependencies in computational communication science with actor – network theory. *International Journal of Communication*, *13*, 3955–3975.

Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., & Nürnbergk, C. (2021). Toward a stronger theoretical grounding of computational communication science: How macro frameworks shape our research agendas. *Computational Communication Research 3*(2), 152-179. https://doi.org/10.5117/CCR2021.02.002.WALD

Waldherr, A., Maier, D., Miltner, P., & Günther, E. (2017). Big data, big noise: The challenge of finding issue networks on the web. *Social Science Computer Review*, *35*(4), 427–443. https://doi.org/10.1177/0894439317690337

Weber, E. P., & Khademian, A. M. (2008). Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. *Public Administration Review*, *68*(2). https://doi.org/10.1111/j.1540-6210.2007.00866.x