Harnessing the Computational and Social Sciences to Solve Critical Societal Problems

Elizabeth Mynatt (Georgia Tech, co-chair)
Duncan Watts (University of Pennsylvania, co-chair)

Nadya Bliss (Arizona State University)
Alondra Nelson (Social Science Research Council & Institute for Advanced Study)
Willie Pearson (Georgia Tech)
Rob Rutenbar (University of Pittsburgh)

National Science Foundation
WHERE DISCOVERIES BEGIN

The NSF CISE Advisory Committee strongly endorses the findings from the round table report entitled "Harnessing the Computational and Social Sciences to Solve Critical Societal Problems," co-chaired by Elizabeth Mynatt (Georgia Institute of Technology, CISE Advisory Committee member) and Duncan Watts (University of Pennsylvania, SBE Advisory Committee member) and funded jointly by CISE and SBE. Starting from a first-ever joint meeting of both CISE and SBE Advisory Committees in late 2019, a joint round table meeting (held virtually in May 2020 due to COVID), brought together a vibrant and diverse community of researchers across the breadth of the CISE and SBE communities, and highlighted a range of joint research themes and cross-cutting challenges. This round table report captures and crystalizes these themes and challenges. A follow-on joint meeting of the CISE and SBE Advisory Committees in December 2020 confirmed the high level of ongoing interest in joint collaboration on these important research topics.

Magdalena Balazinska and Rob Rutenbar,
CISE Advisory Committee co-chairs,
on behalf of the CISE Advisory Committee

CONTENTS PAGE

1.0

BACKGROUND

It is increasingly apparent that many of the systems on which our society depends for its health, prosperity, and security are neither purely social nor purely computational ones. Rather, they are *socio-technical systems*. Workplace relationships, media markets, health delivery systems, and criminal justice organizations are all increasingly characterized by a complex mixture of human actors and institutions on the one hand, and digital platforms and algorithms on the other hand. Efforts to design, manage, audit, and ultimately improve these systems to the benefit of society therefore lie at the intersection of the computational sciences and the social-behavioral sciences.

1.0 BACKGROUND

This intersection has been cast into sharp relief by the COVID 19 pandemic, which--in addition to presenting humanity with an almost unprecedented global public health crisis--has exposed innumerable connections between interpersonal interaction and almost every other element of society. Professional meetings and conferences that were once held in-person are now routinely virtual. Working from home has partially or wholly replaced commuting to the office for millions of workers. Hundreds of thousands of school and university classrooms are now remote, transforming the educational experiences for millions of students and their teachers. Food and package delivery services have gained dramatic market share at the expense of brick-and-mortar retailers. Subsectors of service industries, such as elements of healthcare, travel/transportation, hospitality, among others, are disproportionately facing financial crises.

The long-term impact of these changes--on the economy, on specific industries, on the future of cities, on society and culture--remains highly uncertain, but it is already clear that technology has played, and will continue to play, a critical enabling role. Had this crisis taken place twenty or even ten years ago--as it absolutely could have--the transition to remote work, education, commerce, and social interaction would have been immeasurably more difficult, likely resulting in even more dire outcomes than we have witnessed. Looking forward, technology is also likely to play an important role in our collective response to future pandemics--enabling, for example, real-time detection and contact tracing--with all the attendant concerns about individual privacy and civil liberties.

And the pandemic is not the only event this year that has highlighted the critical effects of technology on society and vice-versa. In the domain of racial justice and disparities, repeated incidents of police violence against people of color--compellingly documented by near-ubiquitous cell phones equipped with video cameras--have sparked waves of outrage and calls for fundamental change to police departments and practices. In the domain of media and democracy, digital technologies such as internet streaming, social media, mobile phones have fundamentally disrupted both the production and consumption of information--and misinformation--with profound consequences for public opinion, political polarization, and trust in institutions. And in the domain of the workplace, the increasing encroachment of artificial intelligence (AI) and other forms of automation is creating exciting new markets for consumers and businesses but also daunting challenges for workers and regulators.

How should the scientific community respond to the increasing relevance of socio-technical systems to every facet of life? How can the research community advance basic science and understanding in a world where the traditional boundaries between computer science and social science disciplines are increasingly unhelpful?

Motivated by these concerns, a joint working group, comprising four members of the NSF CISE and SBE advisory committees and two outside members, convened a one day "virtual roundtable" in May 2020 on the topic of critical societal problems at the intersection of the computational and social sciences.

In this report, we summarize the presentations and discussions that took place during the roundtable. We draw out major themes that came up throughout the day regarding: (a) the characteristics of impactful collaborative research in socio-technical systems; and (b) the nature of the barriers to this kind of research that arise from existing funding mechanisms and research cultures. Based on these themes, we identify several opportunities for the community to foster research in socio-technical systems. Although some exciting progress has been taking place organically in recent years, in particular in the area of computational social science, we conclude that transformational change will require significant institutional commitment in the form of new streams of funding, new models of interdisciplinary collaboration among researchers, and a new relationship among academia, government and industry. Finally, we conclude with some anticipated next steps for the joint working group.

2.0

ROUNDTABLE OVERVIEW

The Virtual Roundtable on Harnessing the Computational and Social Sciences to Solve Critical Societal Problems took place Tuesday, May 19, 2020 from ***10:45 AM to 5:00 PM EST***. The goals of the roundtable were to foster and scaffold collaborative scientific research in the computational and social-behavioral-economic spheres, to underscore the deep interdependence of technological and social systems, as well as to explore ideas to improve collaborations between academia and industry around data, science, and society.

This meeting was originally conceived as a face-to-face workshop scheduled for early May, 2020. As the pandemic unfolded in the US, planning rapidly pivoted to a virtual event. While unanticipated benefits included a diverse and accomplished set of attendees, notable disadvantages included reduced cross-cutting discussions originally planned as breakout groups. Also the economic uncertainty accompanying the pandemic in the spring reduced industry participation, for example in desired sectors such as healthcare and journalism. Nevertheless, the high quality of the invited presentations spurred robust discussion through Q&A sessions moderated by session facilitators alongside active text-based chat across multiple channels.

2.1 AGENDA

In order to anchor our discussions in substantive societal problems, the roundtable was structured around three thematic sessions (see **APPENDIX A** for detailed descriptions of the themes and **TABLE 1** for detailed agenda)

1. Rendering visible, understanding, and ultimately reducing long-standing disparities
2. Improving the trustworthiness of the information ecosystem
3. Empowering and diversifying the technical workforce

We emphasize that these topics were not intended to be exhaustive of the areas in which computational and social scientists could or should collaborate to the benefit of society. Clearly, we could have chosen other topics of importance, including response to COVID 19 and future pandemics, and one possibility for future working group activities would be to expand the substantive focus to these areas. Nonetheless, there was general agreement that these topics were appropriate candidates for discussion, both in terms of their relevance to society and their intrinsic interdisciplinarity.

By bringing together experts from different backgrounds united by a desire to solve real-world problems, these topic areas also served as useful contexts within which to discuss various cross-cutting challenges such as the need for new research infrastructure, industry-academic partnerships, interdisciplinary collaborations, and training and education programs, that participants saw as necessary to facilitate significant progress independently of their domain (see APPENDIX B for detailed descriptions of the cross-cutting issues). Indeed, many of the themes that arose from the discussions--and consequently many of our recommendations to the NSF--concerned these cross-cutting issues rather than the specific domains around which the discussions were oriented. We therefore suspect that many of these same themes and recommendations would have emerged even with a different set of problem areas and different panelists. We note, finally, that many of the same issues have been raised by a subsequent publication in *Science* [1] on the recent history and possible future of computational social science.

2.1 AGENDA

TABLE 1: AGENDA FOR VIRTUAL ROUNDTABLE Tuesday, May 19, 2020

10:45 AM - 11:00 AM

Welcome and Opening Remarks
Elizabeth Mynatt (Georgia Tech)

Opening Remarks by NSF Officials
Margaret Martonosi (Assistant Director, Computing and Information Science & Engineering Directorate)
Skip Lupia (Assistant Director, Social, Behavioral, and Economics Directorate)

11:00 AM - 12:30 PM

PANEL 1: IMPROVING TRUST IN THE INFORMATION ECOSYSTEM

Moderators: *Nadya Bliss (Arizona State University) & Duncan Watts (University of Pennsylvania)*

Cross-disciplinary Collaboration at the Pew Research Center
Claudia Deane (Pew Research)

Building Infrastructure to Support Research and Development of (More) Trustworthy Information System
Kate Starbird (University of Washington)

What Data Do We Need to Study the Information Ecosystem?
David Lazer (Northeastern University)

Auditing and Investigating Networked Information Platforms
Chris Wiggins (Columbia University)

1:00 PM - 2:30 PM

PANEL 2: RENDERING VISIBLE, UNDERSTANDING, AND REDUCING HISTORICAL DISPARITIES

Moderators: *Elizabeth Mynatt (Georgia Tech) & Alondra Nelson (SSRC & Institute for Advanced Study)*

Data Systems for "Rendering Visible, Understanding, and Reducing Historical Disparities"
David Grusky (Stanford University)

The Mythology of Racial Progress
Jennifer Richeson (Yale University)

Health Informatics and Health Equity: Confronting Longstanding Disparities
Tiffany Veinot (University of Michigan)

Measuring and Reducing Disparities
Suresh Venkatasubramanian (University of Utah)

3:00 PM - 4:30 PM

PANEL 3: EMPOWERING AND DIVERSIFYING THE TECHNICAL WORKFORCE

Moderators: *Willie Pearson (Georgia Tech) & Rob Rutenbar (University of Pittsburgh)*

Using Big Data to Understand the Workforce- How Will ML Transform the Economy
Erik Brynjolfsson (MIT)

Pathways to Computing: Growing & Diversifying the Workforce
Nancy Amato (University of Illinois at Urbana-Champaign)

Structural Challenges and Practical Questions for Empowering and Diversity Technical Work in the "New Economy"
Sharla Alegria (University of Toronto)

The Future Workforce: Human-AI-Robot Teaming
Nancy Cooke (Arizona State University)

4:30 PM - 5:00 PM

Closing Remarks / Next Steps
Erwin Gianchandani (Deputy Assistant Director, Computing and Information Science & Engineering Directorate)
Kellina Craig-Henderson (Deputy Assistant Director, Social, Behavioral, and Economics Directorate)
Elizabeth Mynatt (Georgia Tech)
Duncan Watts (University of Pennsylvania)

2.2 PARTICIPATION

As noted above, an unanticipated benefit of moving from an in-person to virtual meeting was that we could accommodate a larger and more diverse pool of participants than the initial **32** in-person invitees. In total, **72** people joined for at least part of the day, with **85%** of participants staying through the end of the workshop. Participants comprised **47** from Academia, **7** from Industry, **2** from non-profit research organizations, and **16** from NSF. Of the NSF participants, **8** were identified as coming from a CISE background and **8** from SBE. Twenty-three (**23**) participants have been previously funded by CISE and **18** by SBE (see **APPENDIX C** for a full list of participants and their affiliations).

3.0

MAJOR THEMES SPANNING ROUNDTABLE DISCUSSIONS

Here we identify the main themes that arose during roundtable presentations and online discussion. Collectively these themes point to characteristics of needed research in socio-technical systems as opposed to specific topics or research questions.

3.1
NEED FOR MORE AND BETTER DATA

A persistent theme throughout the sessions was the need not only for more data but also better data. For example, in the session on trust in the information ecosystem it was noted that attempts to measure the prevalence of misinformation in media inevitably encounter the difficulty that news-relevant information is produced and distributed via a multiplicity of "channels" including TV, radio, and desktop and mobile internet. Because the data associated with these distinct channels are recorded by many distinct organizations all of which have different rules for storing and sharing data, no one entity has a panoptic view of the "information ecosystem." Researchers are therefore left to gather small and often unrepresentative slices of data that are made available via APIs, web scrapes, or one-off data sharing agreements. A result is that we have thousands of papers about Twitter, not because Twitter is a primary source of information for most Americans but because it is the easiest platform for researchers to study. Relatedly, the measures that researchers are forced to use are also typically determined by the idiosyncrasies of the platform rather than the substantive question of interest. Referring to Twitter again, whereas researchers often care about *exposure* to information, what they can measure are user actions such as retweets and follows, metrics that are highly imperfect proxies for exposure.

Similar issues arose in the other thematic sessions. Studies of economic and social disparities are hindered by a lack of data about individual circumstances and life course trajectories that are sufficiently rich (i.e. the right variables measured the right way) and sufficiently broad (i.e. covering all geographic regions and over extended intervals of time). Studies of bias in policing are hindered by the absence of centralized reporting of police records, and practical difficulties associated with obtaining records from hundreds of separate localities. And studies of workplace dynamics are hindered by an inability to experimentally manipulate potential causal effects, or to measure outcomes of long term interest at the scale of entire industries. Although the details differ across specific contexts, a consistent overarching theme of contemporary, data-driven social science research is that while vast amounts of data is in principle available, the most relevant data remains too difficult to collect, too difficult to access, or too difficult to compile, clean, and organize for research purposes.

One solution to problems of this sort is to invest in a new class of research infrastructure. Rather than many individual researchers investing time and resources compiling small, idiosyncratic, one-off datasets for their personal use, an entire community of researchers could pool resources to create much larger, more systematic, and continuously maintained and updated datasets for their collective use. Not only would the resulting data be higher quality and more durable, it would also facilitate reproducibility and replication of key research results, leading to better, more reliable science.

Research infrastructure of this sort would be expensive by the standards of the social sciences. To illustrate, a single national, mobile panel with the capability to collect both high-frequency device and behavioral data as well as rapid polling, and with enough coverage to support local (e.g. sub-city level) analyses might cost upwards of a hundred million dollars per year--nearly half the current SBE budget. From a larger perspective, however, investments of this sort would generate extraordinary value for money. Because the most expensive component of social science research budgets is human labor, any investment in data acquisition that saves time for future researchers pays dividends in savings. Because a single, well designed panel of this scale and scope would replace hundreds of individual data gathering exercises, the effective cost savings over an entire research community might amount to thousands of person years of labor, and with better outcomes.

Another important point that came up in our sessions was that "infrastructure" is more than data storage and computing cycles; it is also the human labor necessary to make the data accessible and useful to other researchers. It is a truism that "90% of data science is data cleaning," but this fact nonetheless poses a significant barrier to entry to many researchers who lack the necessary skills and time to perform their own data cleaning. Augmenting the collection of large-scale datasets with a team of data engineers and data scientists whose role is to support outside researchers can therefore dramatically increase the utility of the data. It was also noted, however, that research support staff, while commonplace in biomedical research labs, particle accelerators, etc., are rarely funded in SBE programs. While CISE researchers may have more available resources, it was also observed that culturally CISE research programs did not require or incentivize allocating resources for data engineers and the curation of reusable and sharable data sets.

Finally, it was repeatedly noted that the collection and use for research purposes of ever larger amounts of data of an ever expanding scope presents serious ethical questions that have not been satisfactorily resolved, in part because they were not anticipated by existing research ethics frameworks such as the Belmont Report. For example, distinctions between primary and secondary data, between private and public disclosure, between research conducted in federally funded institutions and industry, and between research conducted to advance knowledge vs. to develop products, are all intrinsically blurred by research that relies on data originally collected for other purposes. Likewise, questions of data ownership are increasingly difficult to resolve unambiguously. If research is performed for the public good (e.g. to prevent or mitigate the spread of a pandemic) using data aggregated by multiple private companies from millions of mobile devices owned by individual people who are in most cases not identifiable either to the researchers or to the companies, who "owns" these data, and to whom are the researchers accountable? How should one weigh the value of a diffuse but broad public good against an individual's interest in controlling the uses to which "their" data is put? How should one assign ownership of the knowledge that an unidentified person has taken an action such as clicking on a link or visiting a particular location? What right should people have to opt out of passive data collection that could serve a vital public interest? In light of these and other ambiguities, it was generally agreed that any effort to build new research infrastructure should include explicit support for the consideration of accompanying ethical considerations. Moreover it was noted that because the novelty and diversity of the issues involved would likely require the development of new frameworks, ethical considerations would constitute a research program in their own right, not merely something to be "bolted on" to other research programs.

3.2
NEED FOR PARTNERSHIPS

A second broad theme that emerged repeatedly was the importance of partnerships between researchers and other constituencies, specifically industry, government agencies, and local communities.

In the case of industry, there exists substantial disagreement among the academic community regarding the correct approach. On the one hand, it could be argued that in order to be effective, academic-industry partnerships would need to be true partnerships in which the interests of both sides were to be represented. Such arrangements would stand in contrast with previous calls for industry to share data with academics that have assumed, in effect, a one-way transfer of data with no conditions imposed. A key characteristic of a true partnership is that the platform, and not just the data, would be "on the table" for inspection and recommendation for change. Current one-sided arrangements that focus on data "donations" have not built up the buy-in from platform owners invested in evolving platforms based on socio-technical research insights. On the other hand, many academic researchers--especially in the social sciences--view corporate actors as inherently untrustworthy, and hence worry that any partnerships anchored in mutual benefit would fail to serve the public good. A better approach, in their view, would be to persuade the federal government to compel industry compliance, affording researchers access to necessary data without compromising their agendas. Regardless, it was agreed that the current "system" for accessing industry data is hopelessly nonsystematic, nontransparent, and inequitable; thus, something radically different must be devised.

In the case of community partnerships, it was noted that historically academic researchers had not always lived up to the expectations of their partners, especially with respect to marginalized communities. Future arrangements should therefore be required to specify the benefits of the research to the relevant communities--potentially as part of a proposal's broader impact statement--and should be held to account for these commitments. The need for *sustainable* community partnerships was frequently raised alongside the difficulty of doing so via individual research grants and programs. Although current collaborative programs, such as Smart and Connected Communities, require active investment and participation by community partners, the concern remains about the sustainability of these partnerships outside of 3-4 year grant cycles. One thought was that, just as funding agencies invest in collaborative relationships with corporate partners, similar attention could be invested in nonprofit actors, perhaps also managed regionally for local, regional institutions.

Relatedly, because socio-technical systems may have disparate impacts on different subgroups of society, and may therefore exacerbate or ameliorate existing inequities in ways both intended and unintended, it is critical that the design, implementation, evaluation, and analysis of these systems be performed by a research community that is as diverse as the community that is ultimately affected. Although all dimensions of diversity are relevant, particular mention was made of race, gender, age, and disability status.

3.3 NEED TO CREATE & SUSTAIN MULTIDISCIPLINARY TEAMS

A third recurrent theme was that interdisciplinary research has been the subject of innumerable calls to action over previous decades (see, e.g. [2]) but remains difficult to do in practice. Many reasons for this frustratingly persistent state of affairs were discussed, most obviously that academia is intrinsically siloed into culturally and methodologically distinct disciplines. As a result, interdisciplinary research faces a daunting combination of higher difficulty of attracting funding, lower likelihood of being published in top journals, and less recognition by tenure and promotion committees even once published. For example, computer scientists predominantly publish their work in peer-reviewed conference proceedings with fixed submission deadlines, short formats, and rapid publication schedules, whereas social scientists mostly publish in long-format journals that allow multiple rounds of review and for which the time between initial submission and publication may be measured in years. Social scientists may therefore worry that a CS conference paper will not be perceived by their colleagues as "real," while computer scientists may view the opportunity cost of publishing in social science journals as too high. As a result, neither group may be willing to publish their best work in the other's preferred format.

At a deeper level, truly impactful interdisciplinarity often requires more than simply borrowing ideas or methods from domain X and applying them in domain Y. Collaborators coming from different backgrounds may speak different "languages" and hold different assumptions about which questions are important and what standards of evidence are required to address them. "Taking the problem seriously" therefore requires collaborators on all sides to critically examine each other's assumptions and framings before even finalizing the research design. Existing methods may also have to be modified or replaced in order to deal with problems that arise in applications outside of the domains for which they were originally developed. "Deep" interdisciplinary research can therefore take much longer than research that sits within an established discipline, with no guarantee that the additional work will be recognized by peers who are still overwhelmingly disciplinary.

Exacerbating these problems, interdisciplinary training is rarely standardized, forcing students to piece together their own training programs while also satisfying their own departmental requirements. Along with higher course loads, students embarking on such self-directed programs risk compromising their performance in their "core" program without gaining equivalent compensatory recognition for the extra-curricular activities. These students may also be regarded by their advisors as being less serious than equally talented peers who choose to focus on more standard problems and methods, thereby attracting less support on the job market.

For all these reasons, academics (especially untenured) interested in advancing their careers continue to experience significant disincentives to invest in interdisciplinary research. Moreover, because this adverse incentive structure comprises numerous, mutually reinforcing components--funding agencies, journal editors and reviewers, tenure and promotion committees, peers and mentors--it cannot easily be altered. It is therefore no surprise that previous attempts to advocate for more interdisciplinary work have made little headway. Training programs such as IGERT come and go with modest, and often transient, impact on graduate curricular. Funding programs that mandate multidisciplinary teams require proposers to identify their goals and methods *ex ante*, precluding precisely the kind of "deep" interdisciplinary work that can be transformative. And new fields, when they do emerge, often come to be dominated by a single discipline whose norms and tastes they then come to resemble.

Although our discussions did not produce any novel solutions to these longstanding problems, the general consensus was that, in order to produce transformative research at the intersection of computational and social science, the community would benefit from fresh approaches.

- One such possibility that was suggested by the framing of the roundtable itself was to orient new funding opportunities around applied problems rather than around advancement of understanding per se. Such "use inspired" [3] or "solution oriented" [4] research is already attracting attention in the NSF--especially in framing of convergent research--but it has yet to make much impact on the SBE sciences.
- Another possibility, complementing the first, would be to fund centers rather than projects. Such centers could offer teams of researchers the time and flexibility to engage in deep interdisciplinary work, ideally in the service of use-inspired and use-informed problems. Centers of this sort could also be natural homes for the shared research infrastructure and partnership models described above. By creating opportunities to engage in a style of high-impact, high-visibility research that may not be possible in any other environment, solution-oriented research centers may create sufficiently large upsides to interdisciplinary research as to outweigh the risks.
- Finally, the potential longevity and administrative support functions of centers also make them natural homes for interdisciplinary training programs and effective incubators for long-term change in academic environments. One notion worth considering is the ability to link centers, or create academic networks linked to centers, to better facilitate academic and research exchanges. Again, while such exchanges are common in large research grants, they generally do not occur over a sufficient span of time to instigate cultural change.

3.4
NEED TO ORIENT RESEARCH AROUND SOCIO-TECHNICAL PROBLEMS

A final theme that came up, often in parallel with the previous theme, was the time and effort needed to identify and hone the questions at the heart of multidisciplinary socio-technical research. Discussants expressed a need to "iterate on the questions, not just the solutions" especially in the context of forging multi-disciplinary teams.

One variant of this concern was commonly expressed by computational scientists and engineers who experience pressure to identify promising technical approaches at the outset of a research project. For example, technical approaches commonly follow a "utility" maximizing model oriented to technical metrics whereas computational approaches motivated by "harm centered" models, e.g. disparities, require understanding how these harms occur before committing to a technical path. Identifying cross-disciplinary, socio-technical research questions and approaches requires support for this discovery process.

Another variant was an expressed desire to shift the "mode of production" of academic research somewhat away from its current single minded focus on the publication of peer-reviewed papers and somewhat toward the solution of real-world problems. Although the publication of research findings should continue to be an essential feature of open, transparent, reproducible, and ultimately trustworthy science, it was agreed that the traditional metrics of scientific productivity and impact--i.e. publication and citations counts respectively--are not useful measures of societal impact. If the research community is to unlock the kind of resources and goodwill needed to realize some of the other identified needs--expensive research infrastructure, partnerships with industry and community groups, long-term funding for research centers--it must produce outputs that are valued by constituencies other than academic researchers.

Specific examples of such outputs included: apps and web services that are designed to minimize disparities and machine learning algorithms to reveal hidden labor and biases. More generally, outputs would correspond to actionable knowledge that generalizes to the scale and diversity of real-world systems. For example, how should news publishers and technology platforms adapt their operating procedures, algorithms, and user experience features to combat misinformation, reduce affective polarization, and improve public understanding of critical issues? How should public policy more precisely and effectively address racial and geographic disparities in education, wealth, and health? How should training programs, employment practices, and human resource management policies adapt to the changing technological landscape to support both economic productivity and worker well-being? How can predictive modeling, based on granular human mobility data, combined with widespread rapid testing, contact tracing, and supported isolation, help societies respond more effectively to future pandemics?

Complementing this emphasis on use-inspired research, it was noted that more attention should be paid to long-term evaluation of funded research. As with the research design phase described earlier, evaluation also needs to be a critical and reflective exercise in which both the objective function to be measured and the methods for measuring it are open to interrogation. Indeed, research design and evaluation need to be thought of as an iterative process that plays out over extended time intervals--potentially spanning many years--where each "round" of evaluation informs the design of the next round. Once again, however, support for such long-term "integrated" design-research-evaluation programs would require fundamental rethinking of computing and social science funding practices.

4.0

OPPORTUNITIES

Roundtable participants expressed both optimism and pessimism for creating a more productive and impactful research ecosystem that brings together the socio-technical expertise represented by the SBE and CISE directorates.

Looking forward, a distilled list of features for socio-technical, collaborative research between computational and social sciences includes:

- Socio-technical framing to advance societal goals (e.g. a trusted information ecosystem, reducing disparities, empowering and diversifying the technical workforce);
- Multi-disciplinary teaming including cross-training opportunities for students;
- Opportunity to engage long-term basic science in concert with use-inspired, societal goals;
- Research design that includes opportunities for reflection and longitudinal evaluation;
- Support and resources for managing these multi-disciplinary teams including how to manage different incentive and reward structures (e.g. books v. conferences), budget and funding expectations, and equity in opportunities for participation across fields (e.g. teaching and TA loads).
- Reusable and sustainable research infrastructure including data and partnerships

4.0 OPPORTUNITIES

Here we identify cross-cutting opportunities to cultivate collaborative research in socio-technical systems.

1. ***Socio-technical research centers:***
Socio-technical research centers that bring together computer scientists and social scientists would complement individually funded grants. These centers would create the "time and space" to develop strong multidisciplinary teams, to identify "use-inspired" research questions, and to build needed partnerships. Centers would support critical research infrastructure, e.g. curated data, while hosting training programs and sustaining partnerships. One potential model, akin to Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering, could emphasize building connections to industry stakeholders and curating data for researchers across the "hub and spoke" network. Another model, akin to the AI Institutes, could emphasize multidisciplinary research focused on societal sectors such as journalism, healthcare, education and transportation.

2. ***Sustainable research infrastructure:***
The community would benefit from research infrastructure that could be shared across the computer science and social science research communities. Socio-technical data observatories could aim to curate data across a sector such as news consumption drawn from multiple platforms and providers. Consortia of community partners, perhaps organized regionally, could partner on federal funders' research solicitations, akin to current models with commercial industry. Commercial entities could bear a larger share of the costs.

3. ***Budget items for critical infrastructure:***
The research community would benefit from incentives to include dedicated budgets for critical infrastructure for socio-technical research in research awards. For example, the current program in Smart and Connected Communities, requires community partners to be an integral part of research projects, including financial commitments. Data management plans could include budget details for data engineers and support for sharing data, as appropriate, with the larger research community and beyond the time period of the immediate research grant.

4. ***Mechanisms for Teaming:***
The community would benefit from new mechanisms to foster the creation of multi-disciplinary teams. Planning grants could bring together disparate research expertise in combination with relevant partners. Small grants for nascent teams could kickstart new research collaborations. For these grants, it is important to lower barriers for initial success as cross disciplinary reviewing can be doubly critical.

5. ***Opportunities for Education and Training:***
The community would benefit from new training programs to grow multi-disciplinary exposure and baseline training. Creating long-standing programs, such as a computer science/social science summer school, could foster a new generation of multi-disciplinary networks for cross-trained scientists. A review, perhaps an Academies study, that examines the impact of cross-disciplinary educational programs, such as in Human-Centered Computing and Computational Social Science could inform future educational directions.

6. ***Experiment with the reviewing process:***
The challenges of navigating a multi-disciplinary review process can multiply with the number of disciplines involved, resulting in more conservative, incremental research. A light-weight review process for grants that foster initial teaming, planning and multi-disciplinary discovery might lower the barrier to forming productive collaborations. For larger and more competitive grants, the creation of recurring review panels could enable the reviewers to develop greater trust and knowledge in the respective expertise in the panel.

7. ***Address diversity and disparate impact:***
Cutting across all of the above is the issue of diversity and disparate impact. Whether the objective is to support new types of research centers, new types of research infrastructure, or new types of training programs, diversity should be a priority at all phases, starting with design and then continuing through implementation and evaluation. Correspondingly, when thinking through the consequences of new investments in novel scientific enterprises, consideration should be given to how these enterprises will impact different communities--whether in terms of who benefits or who is harmed--and care should be taken to understand and respond to their potential concerns.

5.0

FUTURE STEPS

Each of these recommendations requires further investigation and collaborative design to bring about concrete implementation. We suggest that a small steering committee recruit leadership to plan and organize future activities such as convene needed workshops, roundtables, and solicitations for input from the relevant communities.

5.0 FUTURE STEPS

1. *Design of long term research centers.* Taking seriously the recommendation to stimulate breakthrough research on socio-technical systems by funding a series of long-term research centers, how should such centers be designed? What models are available, from NSF and other agencies, and how would these models translate to the context of socio-technical systems? Which substantive themes offer the most promise for transformative progress over the next ten years under a research center model? Which constituencies would have to be included in order to make such centers successful? How would they be funded?

2. *Negotiating partnerships between academia and industry.* Arguably the least successful element of the June roundtable was getting participation from the relevant industry partners. In order to even have a discussion about industry needs and concerns, and how they might benefit by partnering with the academic research community, it is necessary to get the relevant people "in the room." Yet while NSF does have close relations with industry in other areas, engagement with major tech or media companies on issues of societal relevance around human-centric data and platforms has proven elusive. Some combination of novel thinking and critical mass of the scientific community would seem to be in order, but precisely how to do this remains unclear.

3. *Big ideas for research infrastructure.* In parallel with industry partnerships, more concrete discussions about particular ideas for shared research infrastructure could yield dividends. For example, it was proposed during the roundtable that a large scale (n100,000) national mobile panel would transform the study both of (mis)information consumption as well as socio-economic disparities. But how would such a panel be designed? How much would it cost? Who would benefit? Would the potential insights be sufficiently transformative from a policy standpoint as to justify the costs? Even a single example such as this would require considerable discussion and analysis, but other examples could also be proposed. How should the community prioritize?

4. *Reconfiguring graduate training at the intersection of computational and social sciences.* While many topics require outside input or support, reconfiguring PhD programs to meet the challenges of research at the intersection of computational and social sciences is something that could conceivably begin today. Yet history demonstrates that efforts to stimulate interdisciplinary training are fraught with challenges. What examples of success do we have to build upon? How can these be scaled up? Or do we need new models altogether?

A. ROUNDTABLE THEMES

Rendering visible, understanding, and reducing historical disparities. Today, citizens of different demographic, geographic, and socio-economic backgrounds experience starkly disparate opportunities across a number of dimensions, including educational achievement, economic security, health, likelihood of incarceration, and longevity. In a number of cases, technological advances appear to be combining with historical legacies to increase these disparities rather than reducing them. Can we identify and frame a research program that would tease out the social impact of new technological developments? Can we identify concrete practices that the public and private sectors can use to increase opportunity for all Americans while reducing socially-unproductive disparities? Can research design and innovation that reduces harmful disparities be incentivized?

Improving trust in the information ecosystem. To function properly democratic societies require their citizens to have reliable access to accurate and trusted information on issues of political, health, and scientific nature. Traditionally, this information has been produced by expert communities, disseminated by a relatively small number of publishing and broadcast organizations, and consumed in professionally curated packages and through interaction with trusted organizations and professions. In the past 30 years, this system of information production and consumption has been profoundly disrupted by new technologies, including the web, social media, and mobile devices. Information about essentially any topic can now be produced by any individual with an internet connection, can be disseminated to audiences of essentially any size without intermediation, and can be consumed in a wide variety of formats (video, podcast, SMS, tweet, etc.) on a wide variety of devices. Complicating matters, the distribution and consumption of content is increasingly a mixture of human and algorithmic decision making, further mediated by digital tools, raising concerns about proliferation of biased, inaccurate, and outright false information. New methods are needed for identifying, classifying, quantifying, and tracking misinformation in all its manifestations, as well as measuring its consequences for public understanding and opinion. Moreover, new partnerships are needed between academia and industry to design, build, and evaluate automated tools for mitigating the harmful effects of misinformation at scale and under real-world conditions.

Empowering the skilled technical workforce. America's skilled technical workforce is critical to continuing American science leadership and to fueling the scientific advances that lead to quality of life improvements. However, by many estimates, a considerable number of today's jobs will soon be eliminated or completely transformed. At the same time, the skills needed for providing many non-economic services and a wide range of highly impactful social interactions are likely to undergo similar change. How can research at the intersection of computer science and social/behavioral science create tangible quality of life improvements through helping industry, government, academia, community organizations, and individuals better understand, and more effectively serve, today's and tomorrow's skilled technical workforce?

B. CROSS CUTTING

In addition to understanding possible research agendas for each of these substantive areas, we also need to focus attention on four cross-cutting challenges:

Partnering with industry around data sharing and implementation. Over the past two decades technology companies have built a dizzying array of digital systems and platforms to facilitate interpersonal communication, social networking, e-commerce, information retrieval, publishing, collaborative work, and many other applications. In addition to transforming large swathes of social and economic life, these technologies have also generated a staggering volume and diversity of data that is of potential interest to social and behavioral scientists. Indeed, the explosion of research that has taken place over this time period in data science and computational social science has been overwhelmingly fueled by this "digital exhaust."

Unfortunately, researchers' access to these data is ad-hoc, unreliable, inequitable, and highly non-transparent. In some cases researchers can use publicly facing programming tools (e.g. APIs) but are then subject to rules and restrictions that are typically determined without consideration of research-specific requirements. In other cases, access can only be granted to employees, requiring that researchers or their students work for the company under non-disclosure agreements. Finally, particular datasets are sometimes made available to individual researchers via social contacts or other one-off arrangements. As a result, the vast majority of researchers have no clear means of accessing the vast majority of data collected by companies. Moreover, research results are often impossible to replicate, either because the original data are unavailable or the conditions under which they were collected have changed.

Increased accessibility and transparency around data sharing could also yield considerable benefits for industry. For medium and small technology businesses that lack their own research labs, collaboration with the academic research community could translate directly to valuable insights that would otherwise remain out of scope. Even for large, established tech companies, increased transparency and external collaboration could improve the quality of their services as well as increase public trust.

Building large-scale, shared data infrastructures for research purposes. Improved access to industry data would dramatically accelerate computational social and behavioral science. On its own, however, it would not be a panacea, for at least two reasons. First, data collected by companies and government agencies are often non-representative or biased in other ways (e.g. because some people generate far more data than others, or because different individuals choose to share different levels of information about themselves). Second, because commercial systems are designed to provide useful services, not to answer scientific questions, they may not collect the data of interest in the first place, or may collect it in ways that are difficult to utilize.

An alternative strategy to partnering with industry, therefore, is for the research community to build its own data infrastructures that are designed specifically to support research. Shared research infrastructure is a familiar concept in the physical, biological, and engineering sciences, encompassing examples such as the Large Hadron Collider, the Laser Interferometer Gravitational-Wave Observatory (LIGO), the Hubble Space Telescope, the Sloan Digital Sky Survey, and the Human Genome Project. In the social sciences, the main examples of shared data infrastructure are long-running surveys such as the General Social Survey (GSS), the American National Elections Studies (ANES), and the Panel Study of Income Dynamics (PSID). However, the scale, complexity, and temporality of digital data are sufficiently different from survey data that whole new designs will be required.

B. CROSS CUTTING

Fostering collaborations between computer scientists and social scientists. In addition to more and better data, rapid progress in computational social and behavioral science will require new models of collaboration between researchers. Currently, computer and social/behavioral researchers utilize distinct publishing models, where the former prioritize publishing in annual, peer-reviewed conferences and the latter in journals. Partly as a consequence, computer scientists tend to publish more frequently and with shorter turnaround times than social and behavioral scientists, where the length between initial submission and eventual publication can easily stretch into years. Computer science publications also tend to have more coauthors than their social/behavioral counterparts, where single authored papers or books are still considered helpful (or in some cases required) for tenure and promotion. Finally, norms governing what is considered a valid contribution vary widely between the two communities, with social/behavioral scientists placing more emphasis on theory-driven explanations and computer scientists valuing accurate predictions and/or working systems. Moreover, the spectrum of social and behavioral researchers includes both quantitative and qualitative researchers and experimentalists, offering both hurdles and new possibilities for cross-disciplinary collaboration.

Collaborations between computer and social/behavioral researchers are therefore complicated by conflicting incentives and inconsistent world views. SBE scientists may worry that conference proceedings with a large number of coauthors will not be valued by their peers, while computer scientists may be unwilling to wait for two to four years to publish work in an unfamiliar journal when the same content could be published within a year in one of their own conferences. Accordingly, one outcome of this collaboration might be the creation of a new open-source journal that combines facets of CS and social science norms as part of the necessary infrastructure of a new way of doing research.

Developing training and educational programs. In recent years a number of Computational Social Science courses (e.g. at **Princeton**, certificates (e.g. **Stanford**), masters programs (e.g. **U Chicago**), and summer training programs (e.g. **Summer Institute on Computational Social Science**, **BIGSSS Computational Social Science Summer School**, **Santa Fe Institute Graduate Workshop on Computational Social Science**) have appeared. More broadly, "CS+X" programs such as those at **University of Illinois** and **Northwestern University** have created opportunities for computer science and other disciplines to interact via workshops, joint faculty hiring, and new undergraduate majors. Notwithstanding these welcome innovations, training at the PhD level remains highly segregated between the social science and computer science communities. Although social science students are increasingly engaging in informal cross training, courses like data science and machine learning are not routinely included in their curricular. Reciprocally, computer science students receive little exposure to SBE relevant topics such as causal inference, research design, and substantive theory. Building on existing efforts, guidelines for designing certificates, masters, and full PhD programs in CSS would be extremely valuable.

C. PARTICIPANT LIST

2020 CISE-SBE Workshop on Harnessing the Computational and Social Sciences to Solve Critical Societal Problems

1.	Sharla Alegria	University of Toronto
2.	Nancy Amato	University of Illinois at Urbana-Champaign
3.	Michael Bailey	Facebook
4.	Paul Baker	Georgia Institute of Technology
5.	Mahzarin Banaji	Harvard University/External Faculty, SFI
6.	Nadya Bliss	Arizona State University, Global Security Initiative
7.	Mic Bowman	Intel Corporation
8.	Steve Breckler	NSF
9.	Meredith Broussard	New York University
10.	Anita Brown-Graham	ncIMPACT, School of Government, University of North Carolina at Chapel Hill
11.	Erik Brynjolfsson	MIT and Stanford
12.	Ceren Budak	University of Michigan
13.	Kathleen Carley	Carnegie Mellon University
14.	Siwei Cheng	New York University
15.	Walter Rance Cleaveland	NSF
16.	Nancy Cooke	Arizona State University
17.	Tressie McMillan Cottom	UNC-Chapel Hill (Incoming)
18.	Kellina Craig-Henderson	NSF
19.	Claudia Deane	Clark Atlanta University
20.	Jon Eisenberg	CSTB
21.	Virginia Eubanks	University at Albany, SUNY
22.	Darleen Fisher	NSF
23.	Rayvon Fouché	Purdue University
24.	Nicholas Fuller	IBM Research
25.	Roy George	Clark Atlanta University
26.	Rayid Ghani	Carnegie Mellon Univiversity
27.	Erwin Gianchandani	NSF
28.	Daniel Goroff	NSF
29.	Mary Gray	Microsoft Research
30.	Robert Groves	Georgetown University
31.	David Grusky	Stanford Center on Poverty and Inequality
32.	Evelynn Hammonds	Harvard University
33.	Natalie Hengstebeck	NSF

C. PARTICIPANT LIST

34.	Mar Hicks	Illinois Institute of Technology
35.	James House	University of Michigan
36.	H V Jagadish	University of Michigan
37.	Elva Jones	Winston-Salem State University
38.	Henry Kautz	NSF
39.	Tracy Kimbrel	NSF
40.	Frankie King	Vanderbilt
41.	David Lazer	Northeastern University
42.	Todd Leen	NSF
43.	Margaret Levi	CASBS@Stanford University
44.	Huan Liu	Arizona State University
45.	Arthur Lupia	NSF
46.	Lena Mamykina	Columbia University
47.	Brandeis Marshall	Spelman College
48.	Margaret Martonosi	NSF
49.	Kathleen McKeown	Columbia University
50.	Beth Mynatt	Georgia Tech
51.	Elizabeth Mynatt	Georgia Tech
52.	Lama Nachman	Intel
53.	Jonathan Corpus Ong	University of Massachusetts - Amherst
54.	Manish Parashar	NSF
55.	Andrea Parker	Georgia Tech
56.	Willie Pearson	Georgia Tech
57.	James Ramming	VMware
58.	Betsy Rajala	Center for Advanced Study in the Behavioral Sciences at Stanford University
59.	Jennifer Richeson	Yale University
60.	Laurel Riek	UC San Diego
61.	Ashutosh Sabharwal	Rice University
62.	Eric Schatzberg	Georgia Institute of Technology
63.	Marc M. Sebrechts	NSF
64.	Kimberlee Shauman	University of California, Davis
65.	Katie Siek	Indiana University
66.	Gurdip Singh	NSF
67.	Gabriela Thompson	Intel
68.	Tiffany Veinot	School of Information and School of Public Health, University of Michigan
69.	Suresh Venkatasubramanian	University of Utah
70.	Duncan Watts	University of Pennsylvania
71.	Joe Whitmeyer	NSF
72.	Chris Wiggins	Columbia

BIBLIOGRAPHY

1. Lazer DMJ, Pentland A, Watts DJ, et al (2020) Computational social science: Obstacles and opportunities. Science 369:1060–1062. https://doi.org/10.1126/science.aaz8170

2. Wallerstein I (1996) Open the social sciences: Report of the Gulbenkian Commission on the restructuring of the social sciences. Stanford University Press

3. Stokes DE (1997) Pasteur's Quadrant: Basic Science and Technological Innovation

4. Watts DJ (2017) Should social science be more solution-oriented? Nature Human Behaviour 1:0015