Enhancing and Accelerating Social Science via Automation: Challenges and Opportunities

**Tal Yarkoni¹ Dean Eckles^{2,3} James A. J. Heathers⁴
Margaret C. Levenstein^{5,6,7,8} Paul E. Smaldino⁹ Julia Lane¹⁰**

¹Department of Psychology, College of Liberal Arts, University of Texas at Austin, Austin, Texas, United States of America,

²: Institute for Data Systems and Society, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America,

³Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America,

⁴Bouve College of Health Sciences, Northeastern University, Boston, Massachusetts, United States of America,

⁵Inter-university Consortium for Political and Social Research, Institute for Social Research, University of Michigan, Anne Arbor, Michigan, United States of America,

⁶Michigan Federal Statistics Research Data Center, Institute for Social Research, University of Michigan, Anne Arbor, Michigan, United States of America,

⁷School of Information, University of Michigan, Anne Arbor, Michigan, United States of America,

⁸Survey Research Center, Institute for Social Research, University of Michigan, Anne Arbor, Michigan, United States of America,

⁹Department of Cognitive and Information Sciences, University of California Merced, Merced, California, United States of America,

¹⁰Wagner Graduate School of Public Service, Center for Urban Science and Progress, New York University, New York City, New York, United States of America

Published on: Apr 30, 2021

DOI: <https://doi.org/10.1162/99608f92.df2262f5>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#)

ABSTRACT

Automation plays an increasingly important role in science, but the social sciences have been comparatively slow to take advantage of emerging technologies and methods. In this review, we argue that greater investment in automation would be one of the most effective and cost-effective ways to boost the reliability, validity, and utility of social science research. We identify five core areas ripe for potentially transformative investment, including (1) machine-readable standards, (2) data access platforms, (3) search and discoverability, (4) claim validation, and (5) insight generation. In each case, we review limitations associated with current practices, identify concrete opportunities for improvement via automation, and discuss near-term barriers to progress. We conclude with a discussion of practical and ethical considerations researchers will need to keep in mind when working to enhance and accelerate social science progress via automation.

Keywords: social science, automation, standardization, discoverability, data sharing, data access, validation

Media Summary

The authors of this paper are leading researchers and practitioners in empirical social science. The paper is a result of a 2-day DARPA-sponsored workshop held at the Center for Open Science (COS) in September 2018.

Visit the web version of this article to view interactive content.

Animated Summary

Enhancing and Accelerating Social Science Via Automation

Automation plays a central and rapidly growing role in our daily lives. It has influenced virtually every domain of human activity. However, social sciences lag far behind the natural and biomedical sciences in the use of automation to solve practical and theoretical challenges. Although it is much more difficult to work with data on human beings and organizations, the authors argue that social scientists could nevertheless be harnessing automation to a far greater extent. They identify five areas in which there are enormous opportunities.

The first is standardizing ways of representing data, results, and other scientific outputs, which they argue will generate quick results given that the barrier to entry is relatively low. The second is reducing barriers to data access, which stifle research and reduce replicability. The time for such an effort is particularly opportune given recent legislation. The third is automation for research workflows, data search and discovery, and collaborative tools to build an enormous research graphs. The potential to leverage work done in other fields is substantial, and many of the applications can be adapted relatively easily to the needs of social scientists. The fourth is validation: the automatic detection or correction of methodological errors. The development of a new generation of automated tools would have a potentially transformative effect on our ability to rapidly detect

and correct scientific errors in nascent and previously published work – and can have a large impact on the quality of social science research. Accordingly, this is one natural target for near-term. Finally, the authors see enormous potential to generate automated insight through (1) automated signal discovery, (2) automated meta-analysis, and (3) reasoning systems or inference engines.

The authors conclude with a warning of the ethical issues that must be addressed in any such agenda.

1. Introduction

The social sciences are at a crossroads. The enormous growth of the scientific enterprise, coupled with rapid technological progress, has created opportunities to conduct research at a scale that would have been almost unimaginable a generation or two ago. The rise of cheap computing, connected mobile devices, and social networks with global reach allow researchers to rapidly acquire massive, rich data sets; to routinely fit statistical models that would once have seemed intractably complex; and to probe the way that people think, feel, behave, and interact with one another in ever more naturalistic, fine-grained ways. Yet much of the core infrastructure is manual and ad hoc in nature, threatening the legitimacy and utility of social science research.

We can and must do better. The great challenges of our time are human in nature—terrorism, climate change, global pandemics, misuse of natural resources, and the changing nature of work—and global in scope. They require robust social science to understand the sources and consequences of these challenges in order to inform decision-making. Yet the lack of reproducibility and replicability evident in many fields (Camerer et al., 2018; Dafoe, 2014; J. Ioannidis, 2008; J. P. A. Ioannidis, 2005; Munafò et al., 2017; Open Science Collaboration, 2015) is even more acute in the study of human behavior and in the study of these global problems. It is extremely difficult to share confidential data on human beings, and the rules on such data-sharing vary substantially across countries and legal jurisdictions (Bender et al., 2016; Duşa et al., 2014; Elias, 2018). There is also a serious lack of scientific infrastructure in the social sciences, which often receive much less funding than other sciences—and the problem is particularly acute in many developing countries, further increasing the need for global data sharing and interoperability initiatives (Berman & Crosas, 2020; Moran et al., 2014; Treloar, 2014).

The central argument we advance in this article is that advances in technology—and particularly, in automation—can now change the way in which social science is done, just as it has done in other domains. Social scientists have eagerly adopted new technologies in virtually every area of research—from literature searches to data storage to statistical analysis to dissemination of results.

We identify five core investments that can transform our science:

- 1) *Standardization.* The data, results, and other scientific outputs generated by social scientists are represented in a bewildering array of formats, and this lack of consistency limits their utility. Automation

could be applied to develop standardized, machine-readable representations of social science objects.

2) *Data access*. Researchers who wish to share or use existing data sets routinely face a series of legal, ethical, computational, and logistical barriers. Automated tools, protocols, and infrastructure could reduce these barriers.

3) *Search and discoverability*. The vast majority of social science data and outputs cannot be easily discovered by other researchers even when nominally deposited in the public domain. A new generation of automated search tools could help researchers discover how data are being used, in what research fields, with what methods, with what code and with what findings. And automation can be used to reward researchers who validate the results and contribute additional information about use, fields, methods, code, and findings.

4) *Claim validation*. The social sciences study complex phenomena, and the rate of inferential errors made by researchers is correspondingly high—and often costly. Automation could be applied to develop error-detection systems that automatically identify, and potentially even correct, errors that are time-consuming or difficult for researchers to detect.

5) *Insight generation*. Our individual abilities to generate new social science theories, predictions, and study designs are subject to human cognitive biases and limitations. Automated systems could help generate and test important new insights.

These five challenges are neither independent of one another nor mutually exhaustive; moreover, for the sake of brevity, we largely avoid discussion of technologies and practices that have already been widely adopted by social scientists (e.g., online survey or crowdsourcing platforms like SurveyMonkey and Mechanical Turk). Collectively, however, they span many of the most pressing problems and opportunities facing the social sciences today. We discuss each core challenge in turn, then conclude with a discussion of several practical and ethical considerations that are likely to become increasingly salient as the scope of automation within the social sciences grows over time.

2. Challenge 1: Standardization

Arguably the most pressing short-term barrier to large-scale application of automation in the social sciences is a lack of standards for representing data, results, and other scientific outputs. Almost every automated technology designed to improve social science research will at some point need to perform operations on machine-readable scientific objects. For example, searching for data sets requires standardized query languages and structured metadata; developing algorithms for automated detection of statistical errors requires standardized representations of statistical models or tests; and automatically synthesizing the results of disparate findings meta-analytically is greatly facilitated by the availability of formal ontologies or controlled vocabularies.

Historically, the social sciences have not placed much emphasis on standardization. Many bespoke data formats have been constructed that vary both within and between fields; many fields have standardized on similar, but not always identical, default analysis strategies (e.g., ANOVAs in experimental psychology versus regression in economics). Each journal structures papers in different ways, including differences in citation formats. Electronic copies of papers are stored in different formats; even PDFs contain an ad hoc mixture of genuine text and images of text. Imposing some order on the chaos of bespoke standards and formats that currently pervade social science would enable a wide range of automated applications that have enormous potential to improve the quality and reliability of social science research.

Of course, developing machine-readable standards for the social sciences cannot hope to be a singular effort; we are not suggesting that a one-size-fits-all approach to the social sciences is feasible, or even desirable. Different kinds of research objects, approached from different theoretical and methodological perspectives, will undoubtedly demand fundamentally different kinds of machine-readable representations. Nor are we implying that all (or even most) of social science research will be amenable to standardization—but simply that common standards could be profitably applied much more widely. To convey the scope of the problem and highlight the enormous benefits conferred by potential solutions, we review several areas that are ripe for increased standardization efforts. We deliberately focus on domain-general problems that apply throughout the social sciences, recognizing of course that individual fields will face many other standardization challenges not addressed here.

Visit the web version of this article to view interactive content.

ICPSR 101: What are Metadata (and why are they so important)?

2.1. Data Representation

Perhaps the most common standardization-related challenge social scientists regularly face—often without realizing it—is data representation and organization. Many researchers fail to represent data in a consistent way even from study to study, to say nothing of the enormous variation found between labs or fields. Most social scientists also rely heavily on proprietary data formats (e.g., Excel, SPSS, SAS) that can impose serious barriers to data access. Converging on field-wide or even social science-wide conventions for basic organization of data—including directory structure and file naming, file formats for metadata representation, human-readable documentation, and so on—would have immense benefits. A number of these benefits are illustrated by successful social science data-sharing repositories like the Inter-university Consortium for Political and Social Research (ICPSR), which curates data to standards it has promoted (in terms of file formats, metadata requirements, codebooks and documentation, etc.) (Inter-university Consortium for Political and Social Research, 2009). Curating data to meet ICPSR standards requires significant effort. The large payoff is that it is then much easier for other researchers to access, understand, and make use of one's data set

(Levenstein & Lyle, 2018). Automating the process of data curation, especially if automated curation were built into the research process, could significantly reduce costs to researchers and repositories while increasing data discovery and reuse.

Beyond data sharing, the widespread adoption of a data standard can have many other salutary effects. A good example can be found in neuroimaging research, where the recently introduced Brain Imaging Data Structure (BIDS) standard (K. J. Gorgolewski et al., 2016) has attracted rapid adoption from tool developers and end-users alike. BIDS has given rise to a thriving ecosystem of self-contained ‘BIDS-Apps’ that can be readily applied to any BIDS-compliant data set (K. J. Gorgolewski et al., 2017); to utility toolboxes like PyBIDS that make it much easier to retrieve, query, and manipulate data from a BIDS data set; and to major online data-sharing platforms like OpenNeuro.org that will soon allow users to execute sophisticated, fully reproducible, cloud-based analyses on uploaded BIDS data sets (K. Gorgolewski et al., 2017). Developing and promoting corresponding field-wide standards in the social sciences is likely to have similar benefits.

2.2. Controlled Vocabularies and Ontologies

In fields where many of the entities have been clearly defined (e.g., molecules in chemistry, genomic sequences in genomics, etc.), machine-readable ontologies are often relatively straightforward to develop. Such ontologies delineate which terms or concepts are valid for a domain, what those terms mean, and how they relate to one another; they can greatly facilitate the development of rich, interoperable tooling ecosystems. In the social sciences, by contrast, there is often no clear fact of the matter about what concepts or variables mean, or how they should be used (e.g., what does political polarization mean in the United States, and is it reasonable to operationalize it in terms of voting patterns in congressional elections?). The absence of controlled ontologies impedes theory testing, model building, and data discovery and reuse.

Although some measure of linguistic ambiguity is likely to be unavoidable in the social sciences given the complex subject matter, even modest ontologies that minimally control the vocabulary researchers use would have important benefits (Poldrack & Yarkoni, 2016). For example, in psychology, codifying the knowledge that the concepts of working memory and executive control are closely related would enable semantic search engines to return data sets tagged with either of the two labels even when a query string includes only one. More aspirationally, having a formal ontology for country-level economic and political indicators might one day enable researchers to execute sophisticated analytical queries like ‘estimate the 10-year impact on GDP of switching from an autocratic or colonial system of government to a democratic one, for all African countries, for starting years between 1945 and 1995.’ Some degree of control over terminology is an essential prerequisite for a huge proportion of automation applications, including many we discuss in later sections.

Controlled vocabularies and ontologies already exist in some areas of social science—typically in fairly narrow domains (e.g., Library of Congress subject headings in library science, or part-of-speech tagging in linguistics; the National Library of Medicine’s Common Data Elements cover some social and behavioral domains). A

number of efforts create standards that span the full range of the social sciences; for example, the Data Documentation Initiative (DDI) provides a machine-actionable generic specification for describing surveys and other observational data, and is actively used by many statistical agencies and academic data repositories (Vardigan et al., 2008). However, DDI was developed for observational data; there is a need for extensions or alternative specifications that target or encompass other kinds of data, as acknowledged in the recent release of DDI-CDI (Cross Domain Integration). An open challenge for social scientists and informaticians is to develop abstractions that hide much of the complexity in specific subdomains of social science in favor of standardized entities common to multiple subfields or even entire disciplines.

Visit the web version of this article to view interactive content.

ICPSR 101: What is Data Curation?

2.3. Harmonization Across Standards and Fields

Our discussion so far may make it sound as though the intended end goal for each of the above areas is a single standard that researchers within a given field agree to abide by. However, complete consensus is neither likely nor desirable; there will undoubtedly be competing technical solutions to each of the problems identified above. Rather than encouraging a winner-take-all outcome, an efficient way to ensure commensurability while encouraging innovation is to develop meta-standards that support common access to different standards. In software development, the need to interoperate with multiple tools that have different interfaces is often solved by introducing a higher level abstraction that exposes a uniform user-facing interface while internally querying multiple services as needed. The same kind of approach will likely be necessary to reconcile disparate standards and resources in the social sciences. One example of an existing effort along these lines is the SHARE initiative (share-research.org), which provides a unified interface to humanities and social science resource outputs aggregated from over 100 disparate sources.

2.4. Outlook

Standardization is a prerequisite for almost every potential application of automation to social science research. Consequently, standardization initiatives offer perhaps the most attractive target for short-term resource investment if one's goal is to catalyze development of a rich ecosystem of interoperable, automated social science applications. It helps that the barrier to entry is relatively low: The relative lack of formal standards in the social sciences means that even small groups of motivated, skilled researchers currently have the potential to introduce new standards that could ultimately end up shaping the trajectory of broad research domains. Community consensus is critical here, as the proliferation of different standards could reinforce silos and undermine the transdisciplinary research critical to transformative social science.

Naturally, there are also a number of practical hurdles. First, because the problem space is massively heterogeneous, solutions will have to be correspondingly varied; this is particularly clear in the case of controlled vocabularies, where the scope for social science–wide standards is probably fairly limited (e.g., to basic statistical methods and key variable classes). Second, when developing new standards, there is often a strong tension between simplicity and comprehensiveness: While it is easier to learn and implement a simple standard than a complex one, simple standards necessarily leave out many important use cases. Third, diversity and innovation must be carefully balanced against consistency and commensurability: While some measure of competition between standards is desirable, researchers should attempt to extend and improve on existing standards before introducing new ones.

3. Challenge 2: Access to Data

A second set of challenges routinely faced by researchers in all areas of social science concerns access to data on human subjects—or, rather, the lack thereof. Some barriers to data access are institutional in nature; for example, in the use of archival administrative records in social science research. “The current legal environment for data collection, protection, and sharing lacks consistency, leading to confusion and inefficiency among departments, external researchers, and other members of the evidence-building community” and “formal data access agreements (e.g., Memoranda of Understanding or MOUs) between two or more agencies can take years to develop” (Commission on Evidence-Based Policymaking, 2017). Similarly, many institutional review boards (IRBs) are focused primarily on evaluation of high-risk clinical trials and invasive biomedical studies, and lack expertise in lower risk social science research—often leading to unnecessary restrictions on data access and data sharing (National Research Council, 2014). In other cases, barriers to widespread data access are cultural in nature, or reflect misaligned individual-level incentives. For example, a recent survey of 600 psychologists suggested that many researchers view data sharing as discretionary, unusual, or time-consuming (Houtkoop et al., 2018).

As a result, the rapid rate of novel data generation in social science is undermined by low rates of data sharing, and even lower rates of data reuse (Wallis et al., 2013). For example, nearly half of articles published in the flagship economics journal *American Economic Review* received an exemption from the journal’s mandatory data-sharing policy between 2005 and 2016 (Christensen & Miguel, 2018), making data reanalysis effectively impossible. Although massive investments have enabled the sharing and reuse of open data, the hundreds of millions of dollars that statistical agencies, universities, research organizations, and philanthropic foundations have spent to make better use of sensitive data have had limited impact on social science research. The reasons are well understood. As noted above, there are fundamental structural problems: The startup costs to access and use confidential data are daunting, and the rewards to individuals are too low. In prosaic terms, the data plumbing needs to be installed before the evidence house is built, and an investment in plumbing has been lacking.

Given constraints on resources, organizations wanting to promote better access to data will want to focus those resources on data sources with the widest and most valuable research utility. Unfortunately, organizations often don't know (and can't report on) who is accessing what data, in what settings, for what purposes, and with what results. On the legal side, organizations face a range of security, auditing, and compliance requirements (GDPR, HIPAA, FERPA, FedRAMP, Title 13, Title 26, SOX, GLBA, etc.). On the technical side, complex and costly computational infrastructure must be built and maintained. All of this is compounded further when research questions involve sensitive data sets from multiple providers, each subject to their own constraints. Few organizations with research capacity are presently able to tackle these legal, technical, and logistical barriers.

The main problems organizations face can be summarized as ensuring that the five 'safes' are addressed: safe people, working on safe projects, in safe settings, safe (deidentified) data, and safe (disclosure proofed) outputs are released.¹ The core challenge is thus to develop automated systems that facilitate much wider access to data in the face of legal, ethical, logistical, and technical barriers, while simultaneously respecting those barriers when they are set up for socially important reasons. We consider a number of representative applications here, including infrastructure for data dissemination and access control, platforms for automated, adaptive training, and tools for automated data extraction.

3.1. Data Infrastructure

The nature of research using data on human subjects has fundamentally changed as the surge of new types of data have become available, but infrastructure for researchers in fields dealing with such data has not. New types of data management and programming capabilities are needed, as are new ways of summarizing, describing, and analyzing massive and unstructured data sets (Einav & Levin, 2014; Lane, 2016). Such advances revolutionized much of science, ranging from life sciences to astrophysics. This did not happen by chance, but as a result of careful, deliberate investment in infrastructure and agreement on data-sharing and management principles—the Bermuda Accord, in the case of the Human Genome Project (Collins et al., 2003), and the Sloan Digital Sky Survey, in the case of astrophysics (York et al., 2000). Similar major investment should be made in social science disciplines if research is to advance on core human issues like child health and learning, the results of investment in research, and the health and well-being of cities. That investment should include a focus on both technical and workforce capacity.

The data revolution has not only produced massive new types of data on human subjects, but has also changed the way data sets are typically collected and disseminated (Jarmin & O'Hara, 2016; Lane, 2016), as well as the skills needed to access and use them. Researchers using data on human subjects need to collect, manage, and use diverse data in new ways. There are many technical challenges that can be automated to reduce costs and promote transparent data stewardship (G. King, 2014). Automation could speed up and standardize the way in which disparate data sets can be ingested, their provenance determined, and metadata documented. Automation could ensure that researchers can easily query data sets to know what data are available and how they can be

used. Automation could similarly ensure that the workflows associated with data management and linkage can be traced and replicated. Automated processes could track access and use so that data stewards could be sure that the bureaucratic and legal issues are addressed. Finally, automated rather than ad hoc procedures could be instituted to ensure that the data are securely housed, and privacy and confidentiality protected.

In connection with the last of these goals, a good deal of attention has been paid to the promise of *differential privacy* approaches to privacy and confidentiality (Dwork, 2008). The core idea here is to introduce a formal policy that guarantees a prespecified level of privacy by modifying query results before they're sent back to the user—most commonly by injecting noise with carefully tailored properties into the data. Although differential privacy has appealing theoretical benefits, its practical utility for many social science applications remains unclear. Among other concerns, differential privacy algorithms require strong formal assumptions about the level of acceptable privacy risk. In most social science contexts, no consensus exists about what constitutes acceptable risk, and society has no process for achieving such a consensus. In some cases, the levels of noise infusion necessary to achieve target levels of privacy seriously reduce data quality; the impact on data quality may not be apparent to a data user for whom the algorithmic data privacy provisions are a black box (for reviews, see Bambauer et al., 2014; Dankar & El Emam, 2013; Lee & Clifton, 2011; Ruggles et al., 2019). The most important use cases for differential privacy at the current time are in producing public use data products for which ease of access is more important than data accuracy (e.g., Google searches trends). Many critical social science innovations require higher quality data, for which another solution to the privacy-utility tradeoff is necessary.

3.2. Training Platforms

Training can also be delivered in a less ad hoc manner. Training is critical. While the new types of data have enormous appeal in their potential to describe economic and social activity, the data science and numerical computing skills needed to take advantage of such data are still taught predominantly in computer science or statistics departments. As a result, while there are many courses in machine learning, database management, or search optimization, the content of those curricula is typically not focused on how to use these skills to contribute to the scientific analysis of social problems. In addition, most such programs are targeted at traditional graduate students, and are part of a continuous course of study. There is a great need for graduate training for all kinds of students—including those in government agencies and the private sector—who need to understand how to use data science tools as part of their regular employment. It will be important to train these students to use new models and tools while grounding the approach in fundamental statistical concepts like population frames, sampling, and valid inference. This will require combining foundational social science research principles with current analytic and computing skills, ideally grounded in the study of real-world social and economic problems.

The last few years have seen a growing response to this need. Many universities have degree or certification programs dedicated to better training in computational methods in the social sciences (Lazer 2020). Moreover,

there are increasing numbers of short courses on computational methods for both social scientists and industry professionals, including Data Carpentry, Software Carpentry, Applied Data Analytics at the Coleridge Initiative, and the Graduate Workshop in Computational Social Science at the Santa Fe Institute. Such programs are highly promising, and in our view represent the future of training for quantitative social science, but the methods they promote are still underrepresented in most social science training programs.

Delivering training programs at scale requires automation as well. While online classes have been developed to use open data, there are only limited in-person training sessions for confidential microdata on human subjects (integrated online and in-person training with confidential microdata are described in Kreuter et al., 2019; Weinberg et al., 2017). Levenstein et al (2018) document the lack of consensus within the data community about what such training should include.

3.3. Automated Data Extraction

The digitization of most economic transactions has fundamentally changed the kind of information available on individuals, households, and businesses. Passively generated records of day-to-day transactions and interpersonal interactions offer the opportunity to improve the measurement of key social and economic indicators while drastically reducing the effort required to acquire such data. In addition, declining response rates and increasing costs of household and business surveys provide additional incentives to explore new source data for economic and social measurement. The current use of commercial organizations' data for social science research is often ad hoc; a framework is needed to move from these ad hoc arrangements to a systematic, institutionalized approach. Successful automation of the extraction of economic and social content from transaction and other nondesigned data has the potential to transform social science research in multiple arenas by creating much more frequent and timely data—'nowcasting' (Antenucci et al., 2013; Giannone et al., 2008)—of social and economic phenomena that are now only observable with long lags.

Automating extraction of this social and economic content requires collaboration between data and social scientists, particularly given the enormity of the cross-sectional data (millions or billions of transactions at any point in time), and as contrasted with the sparsity of the time-series data relative to social time (e.g., time series of Uber trips are very dense over the span of time that Uber has existed, but very short relative to the business cycle or the pace of technological change in automobiles). This requires models that integrate data analysis with domain-specific knowledge to leverage the available data for robust social and economic measurement. It also requires building collaborative tools that facilitate these synergies, as well as developing sustainable mechanisms for maintaining such tools over time in the face of changes to the data being processed (e.g., when the structure, format, or content of the user data returned by Facebook's API changes).

3.4. Outlook

There is general recognition that it is important to build infrastructures to improve access to social science data and promote collaborative analysis. Indeed, Card (2010) call for competition among federal agencies to

provide secure access and reward performance. A new opportunity has arisen with the call for a National Secure Data Service (Commission on Evidence-Based Policymaking, 2017), and the first steps have been implemented with the passage of the Foundations for Evidence-Based Policymaking Act of 2018. Collaboration between statistical agencies and the broader research community is necessary to reap the potential benefits of these steps and create the basis for building a national framework for data access using modern automated technologies.

Visit the web version of this article to view interactive content.

Podcast Interview

Democratizing Our Data with Julia Lane

4. Challenge 3: Search and Discoverability

Good scholarship involves effective use of existing data, results, and ideas; it requires the ability to dredge the vast sea of publicly available research objects for the small subset that one deems both relevant and reliable enough to build upon in one's own work. Unfortunately, this is often quite challenging. Scientific research is overwhelmingly reported in journal articles and conference proceedings that have access restrictions and insufficient metadata for reliable semantic search; consequently, researchers who fail to use exactly the right search terms can easily miss out on entire relevant literatures. Scholarly search engines like Google Scholar regularly miscategorize journal article and book chapters—and do not indicate what data were used in the articles and chapters, or where it can be found. Papers reporting results that have failed to replicate or even undergone retraction continue to be cited for years (Bar-Ilan & Halevi, 2018; Gewin, 2014; Greenberg, 2009; Korpela, 2010), largely because published replications are not readily linked to the original research and to one another. Formal machine-readable connections between research findings are practically nonexistent. And even when researchers manage to overcome such hurdles and successfully identify relevant potential resources, they then typically still have to determine whether or not they should *trust* those resources.

Vannevar Bush foreshadowed these issues more than 60 years ago:

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. ... Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it; and this sort of catastrophe is undoubtedly being repeated all about us, as truly significant attainments become lost in the mass of the inconsequential. (Bush et al., 1945)

A salient recent illustration of this challenge can be found in the scientific community's response to the global COVID-19 pandemic that began in 2020. In the first few months of 2020, biomedical and social scientists published thousands of articles on COVID-19 (Brainard, 2020; Meng, 2020)—many via rapid publication channels putatively justified by their potential relevance to policy decisions with life-and-death implications. Unfortunately, the sheer number of such articles, coupled with highly variable quality-control procedures, made it extremely difficult for policymakers and scientists alike to separate signal from noise. Social media channels were replete with boom-and-bust news stories often driven by shoddy research that arguably should never have attracted widespread attention in the first place (e.g., the widely available drug hydroxychloroquine was initially lauded as a helpful adjunct treatment for COVID-19 on the basis of a handful of poorly controlled studies, before better designed investigations demonstrated it was ineffective and possibly even harmful).

Resolving such difficulties consumes an enormous amount of time and energy for many social scientists. Automated tools and services could greatly facilitate the process—often by passively capitalizing on the accumulated labor of one's extended research community. We break the problem up into three separate subchallenges focused, respectively, on (1) fully reproducible, standards-compliant analysis workflows that generate comprehensive metadata at the same time as the primary scientific objects; (2) machine learning tools for large-scale, efficient annotation and retrieval of existing research objects; and (3) collaborative filtering tools that can help evaluate, curate, and synthesize the enormous emergent research graph.

4.1. Reproducible Workflows

Common standards of the kind described in Challenge 1 are a minimal prerequisite for constructing massive, machine-readable research graphs that formally link data sets, papers, analyses, and ideas, but they are by no means sufficient. Even if a consensus existed about what standards should be used to annotate research objects, we would still need tools and platforms that can facilitate their actual application to the research literature. Part of this problem is arguably already capably addressed by existing technologies that dramatically improve scientific reproducibility and provenance tracking. For example, the rapid uptake of interactive notebooks (e.g., Project Jupyter; jupyter.org) that facilitate integrated presentation of data, analysis code, results, figures, and text has made it considerably easier for authors to construct, share, and interact with reproducible analysis workflows (Kluyver et al., 2016; Shen, 2014). Containerization technologies like Docker take reproducibility even further by ensuring that all software and data dependencies can be bundled into a single self-encapsulated container that can (in principle) be executed on most major platforms (Boettiger, 2015). Building communally adopted standards directly into containerized applications could, in principle, give rise to entire ecosystems of containerized, interoperable applications that annotate and curate social science resources with minimal effort on the part of researchers (cf. efforts like Galaxy and BIDS-Apps in other fields, Afgan et al., 2018; K. J. Gorgolewski et al., 2017).

An admitted limitation of the above vision is that, at present, many social scientists lack the technical skills needed to take full advantage of existing (mostly programmatically driven) technologies. While we believe the

ideal long-term solution to this problem is for social science graduate programs to greatly increase emphasis on scientific computing skills, a complementary and perhaps more realistic near-term approach is to focus on designing platforms that lower barriers to usage by hiding the full complexity of a modern analysis workflow behind simple command-line tools and intuitive graphical interfaces. For example, developing free web-based analogs of widely used statistical packages like SPSS and SAS would enable passive annotation and curation of data sets and results with little or no additional effort on the part of users (though access to the generated resources would, of course, remain under users' control). An incipient but illustrative example can be found in the Coleridge Initiative (coleridgeinitiative.org), which seeks to create a highly contextual, engaging computational platform for researchers working with sensitive data sets. The long-term goal is to build user interfaces that present rich context to users about the data sets as they work in secure environments, while also incentivizing users to participate in the creation of additional rich context. These interfaces act to gather metadata that provide information about who else has used the data, for what purpose, and how others users have accessed and analyzed data in their research work.

Visit the web version of this article to view interactive content.

Video

NSF "Rich Context" Data Scorecard

4.2. Annotation and Knowledge Representation

While generating comprehensive metadata at the same time as the corresponding primary research objects is an ideal worth striving for, it is unrealistic to expect researchers to rapidly converge on a single set of best practices—and such an outcome would still do nothing for the vast sea of research objects created in the days before Jupyter notebooks and Docker containers. A prospective, forward-looking approach should therefore be complemented with tools designed to retroactively annotate the existing literature. Rather than approaching this as a social science-specific problem, we suggest that social scientists look to, and adapt methods from, machine learning and computational linguistics—fields that are rapidly generating a wide range of models suitable for use in annotation and search applications. In particular, recent advances in embedding techniques (Mikolov et al., 2013; Zhang et al., 2016) have led to significant improvements in semantic representation of complex texts—including scholarly papers—using low-dimensional spaces that allow efficient and accurate information retrieval (Kusner et al., 2015; Ye et al., 2016). A number of promising applications have already emerged in biomedical fields (e.g., Habibi et al., 2017; Kim et al., 2017), and there is similar potential in the social sciences. We envision the future development of large-scale, continuously learning research engines that would automatically encode every research object in a common, domain-specific latent semantic space—a development that would greatly facilitate a wide range of applications, including automated categorization and annotation, semantic search, and personalized recommendations.

4.3. Collaborative Filtering

The problem of extracting relevant, high-quality information from a vast sea of background noise is not limited to the social sciences, or even to academic scholarship. The need to encourage prosocial participation in a community structure is a widespread feature of collective action problems (Nyborg et al., 2016; Ostrom, 1990; Poteete et al., 2010). In many domains, people have tackled this challenge by building collaborative online platforms that encourage users to both generate and curate an enormous amount of high-quality content. Examples range from domain-specific expert communities like Stack Overflow and TripAdvisor—where answering other users' questions is the explicit purpose of the platform—to general infrastructure platforms like GitHub that have other primary goals (e.g., to support open-source software development), but also provide excellent tools for discovering, filtering, and using platform contents.

Online platforms that effectively solve the information curation challenge have a number of common features. They typically provide clear visual feedback—and a sense that one is contributing to a communal enterprise—following even minor user actions, and they reward increased user engagement with new abilities, community feedback, or greater visibility. For example, in data.world, user onboarding happens immediately from sign-up, and super-users are highly visible from the homepage. On StackOverflow, users gain a range of editorial powers (e.g., the ability to edit others' answers) as their user score increases. And on GitHub, the authors of highly starred repositories gain in reputation and exposure, often leading to new career opportunities.

Arguably the most general common feature of such systems is that they work hard to align individual and communal incentives—that is, to ensure that what's good for an individual user is also good for the entire community. When such an alignment is absent—as is arguably often the case in social science research—individuals may continue to behave in ways that harm the collective endeavor (Liu et al., 2014; Smaldino & McElreath, 2016). Minimizing the transaction costs to participation and ensuring that such participation is appropriately rewarded is a critical part of the solution. Adapting successful modes of online collaboration to the needs of social scientists could help improve the quality and efficiency of multiple aspects of the typical research workflow, ranging from literature and data set discovery to postpublication peer review (Kriegeskorte et al., 2012; Yarkoni, 2012).

4.4. Outlook

The outlook for applications that boost search and discoverability via automation is extremely positive in our view—in large part because many of the applications discussed here are domain-general, and tools and platforms introduced in other fields are likely to be adapted relatively easily to the needs of social scientists. Given the rapid technical progress researchers are making in fields like machine learning, we think the primary focus of social science researchers and funding agencies intent on improving search and discoverability should be on adapting and adopting existing techniques rather than on de novo methods development. Collaboration between social and data scientists will have a large payoff as each learns from the other. Challenges include

bridging disciplinary differences in terminology and publication norms. These can be addressed by a critical mass of forward-looking institutions or funding agencies who commit dedicated and substantial resources to support these collaborative objectives.

5. Challenge 4: Validation

Human-caused methodological error is pervasive in social science, but it is labor-intensive to detect errors. A fertile and largely unexplored application ground for increased automation is the automatic detection or correction of methodological errors. Automation of correctness checking is a very new line of work within social science, but the recent success of the Statcheck framework (Epskamp & Nuijten, 2014)—a tool for detection of basic statistical errors that has been automatically applied to thousands of published articles (Baker, 2016)—suggests substantial growth potential. Four kinds of automation that we believe would be especially fruitful are: (1) automated evaluation of data quality; (2) automatic enumeration of statistical assumptions that are implicit in analyses; (3) automatic sensitivity analysis to assess the range of assumptions that support a result; and (4) automatic creation of checklists for researchers to verify their in-progress work.

While all of the above quality assurance steps are possible today, most currently require additional human labor that is time-intensive and is itself error-prone. As an example, consider Young’s work assessing the validity of parametric modeling assumptions in economics papers, which exposed many results for which nonparametric techniques that make fewer assumptions would have led to substantially different inferences (Young, 2015). While this type of effort is extremely valuable, it requires a direct time commitment from a small pool of experts whose time is scarce relative to the amount of work that could be reviewed. The development of a new generation of automated tools would have a potentially transformative effect on our ability to rapidly detect and correct scientific errors in nascent and previously published work.

5.1. Quality Control

Most scientific claims are only as good as the empirical data they derive from. In domains where researchers work with extremely large, highly structured data sets (e.g., neuroimaging, statistical genetics, or astrophysics), automated quality control (QC) checks and correction procedures are an integral part of many workflows. By contrast, automated assessment of data quality is relatively rare in the social sciences—despite the rich tradition of measurement and methodology these fields often feature. While many researchers already rely heavily on programmatic tools to visualize and inspect data or compute diagnostic quantities, such procedures overwhelmingly require explicit, deliberate effort, offering neither scalability nor any guarantee of reliability.

To some extent, the lack of automated QC tools in the social sciences is unavoidable given the immense heterogeneity of social science data (and associated research questions). However, within specific niches—for example, online surveys involving widely used measures—there is considerable potential for development and deployment of automated methods. For example, detection of outliers in survey data, or anomalies and change points in time series data, are widely studied problems amenable to some degree of automation even in data

sets that lack machine-readable annotations. And once standards of the kind described in Challenge 1 are adopted, the range of feasible QC applications expands considerably. For example, given multiple data sets that have several standardized columns in common, automatically inspecting the covariances between variables could rapidly identify anomalous—or at least, unusual—data sets (e.g., if self-reported age shows a radically different relationship with income and wealth variables in one data set as compared with all others). Centralized databases could be developed that allow researchers to easily compare their own newly acquired data against a continually expanding reference set—often without having to upload or share potentially sensitive data. For example, how do the moments of the distribution of scores on a standard depression measure compare to those extracted from hundreds of previous studies that used the same measure? Such applications could greatly increase the likelihood of detecting errors or unexpected features of certain kinds of social science data, improving the reliability and validity of downstream analyses and increasing general trust in the literature.

5.2. Enumeration of Statistical Assumptions

Many papers have investigated the importance of statistical assumptions by manually reanalyzing large bodies of existing work. G. King et al. (1998) described the importance of assumptions about the safety of ignoring missing values, and reanalyzed several papers to demonstrate the practical impact of this often implicit assumption. We believe that this assumption is now amenable to automatic evaluation using tools commonplace in software engineering. For example, it is common to ‘lint’ computer programs to check for known patterns of undesirable programming practices by having one computer program read other computer programs and catch errors that can be detected based on obvious textual features (e.g., the use of functions that have been deprecated or are known to be inefficient). Even more sophisticated analyses are possible using tools based on static program analysis—a field of research that investigates the properties of programs that can be learned without executing programs (e.g., Chess & McGraw, 2004; Zheng et al., 2006). NASA has used static analysis tools to catch errors in the Mars Rover and other mission-critical systems (Brat & Klemm, 2003); Boeing uses static analysis to verify the safety of airplane designs. Such approaches could be applied to catch simple concerns like the frequency of ignoring missing data or the exclusion of intercepts in regression models. The output of such systems could be a standardized representation of statistical assumptions.

5.3. Robustness/Sensitivity Analysis

Empirical research frequently involves ‘robustness checks’ where researchers evaluate how sensitive their results are to alternative analysis specifications (Steege et al., 2016). Less common are more formal sensitivity analyses that consider, for example, the consequences of unobserved confounding for causal inference. The near-complete lack of the latter type of analysis raises serious concerns about the credibility of the published social science literature. To quote economist Charles Manski, “If people only report robustness checks ... that show that their stuff works and they don’t push it to the breakdown point, then that’s deception

basically” (Tamer, 2019, p.32). Automated robustness checks and sensitivity analyses could potentially play a large role in increasing the credibility of—or detecting problems with—reported findings.

As one example, recent work by Schuemie and colleagues has examined the degree of latent confounding of effects in epidemiology by estimating distributions of estimated effects for known nulls (Schuemie et al., 2014). This allows the automatic construction of empirically calibrated tests and confidence intervals for new analyses. Another promising broad direction lies in automated construction of “multiverse analyses” (Steege et al., 2016)—that is, analyses that construct distributions of alternative results by systematically varying data preprocessing choices, selected variables, estimation parameters, and so on. Early efforts in this direction used millions of distinct specifications of statistical models (Sala-i-Martin, 1997), but such approaches have not yet been widely adopted by authors. Development of general-purpose, easy-to-use tools for robustness analysis would likely improve matters. In cases where authors remain hesitant to spontaneously report robustness analyses, automated robustness checks could still be used during the review process, or by postpublication review platforms, to detect potentially fragile results (e.g., in cases where data are publicly available, by automatically evaluating a large range of alternative model specifications that include additional data set columns as covariates).

5.4. Reporting Checklist Verification

In health care settings, the mandatory use of checklists has been shown to substantially improve a variety of clinical outcomes (Hales & Pronovost, 2006; Haynes et al., 2009). There is an analogous movement in science to promote scientific reproducibility and reliability by encouraging or requiring authors to complete reporting checklists when submitting articles for publication (Aczel et al., 2020; LeBel et al., 2013; Nichols et al., 2017). Unfortunately, enforcement of such checklists currently requires human attention, and hence is often lax. Automating the process of verifying that authors have provided required information—and potentially even validating the information itself—could help prevent many errors of omission at the front end of the publication process (i.e., when submitting articles to journals). For example, we can envision automated, domain-general manuscript-checking tools that (probabilistically) detect whether or not authors have included required items like sample size justification, formal model specification, preregistration documents, or links to external data. Building on some of the other assessment applications discussed above, more elaborate versions of such platforms could perform cursory quality assessment over these items—for example, flagging for human attention cases where reported sample sizes deviate substantially from other works in the same area, where the numbers in a preregistration document appear to conflict with those detected in the manuscript body, and so on.

5.5. Outlook

As the success of Statcheck demonstrates, even relatively simple automated validation methods that can be implemented, given present standards and technologies, can have a large impact on the quality of social science

research. Accordingly, one natural target for near-term investment in this area is development of centralized, automated versions of existing quality assessment tools—for example, various QRP and bias detection methods (Brown & Heathers, 2016; Gerber & Malhotra, 2008; Heathers et al., 2018; Simonsohn et al., 2014), reporting checklists, and so on—that are presently deployed manually and sporadically. More sophisticated kinds of validation—for example, the ability to automatically identify errors in formal mathematical models—will likely improve progressively in the coming decades as computer science, statistics, and social science make further advances. Many of the tools we envision here will also have as prerequisite the kinds of standardized representations of research objects that we discussed in Challenge 1.

Going forward, we perceive two main risks in this area. First, automated validation tools may fail to attract widespread adoption if users perceive that the required effort investment outweighs the prospective benefits. Aside from reiterating the importance of wrapping sophisticated programmatic tools in simple, intuitive user interfaces, we believe that the most effective means to ensure widespread adoption of automated validation methods may be to integrate such methods into centralized platforms that researchers already routinely interact with—for example, by having grant proposal and journal submission systems automatically run a suite of automated QC analysis checks on every uploaded document.

Second, as we have already seen in the case of Statcheck, there may be opposition to automated flagging of errors that are perceived (rightly or wrongly) to be invalid (Baker, 2016). In software engineering, for example, linting tools often impose subjective style rules that stir controversy and require social coordination and proper incentives to resolve. We do not pretend that every automated validation tool will produce error-free results that satisfy every researcher; regular calibration and iteration will undoubtedly be required to ensure that such tools continue to perform at a satisfactory level. The initial contribution of such tools may be to provide a limited number of cases suitable for manual analysis before full automation is possible.

6. Challenge 5: Automated insight

Perhaps the most tantalizing opportunities for automation-related advancement of social science lie in the area of scientific discovery and insight generation. While it is not hard to convince most social scientists that automation could one day play a major role in standardizing data, improving search, or detecting errors, the notion that machines might one day exhibit something akin to scientific insight or creativity often elicits considerable skepticism. To be sure, the prospects of such advances are more distant than those discussed elsewhere in this article. Nevertheless, we believe there are several areas of research where automated technologies could plausibly be deployed within the next decade to produce important new scientific insights. Examples include (1) automated signal discovery, (2) automated meta-analysis, and (3) reasoning systems or inference engines.

6.1. Signal Discovery and Hypothesis Generation

The social sciences are increasingly awash in high-dimensional, unstructured data. Unlike most purpose-built, lower dimensional data (e.g., a simple randomized lab experiment, a short survey), it's often unclear how to extract information relevant to basic and applied sciences from such sources. Even many purpose-built data sets are increasingly quite high dimensional (e.g., neuroimaging data or surveys linked to genetic and other biological data). Other data sources (e.g., large collections of images of urban environments) have clear relevance to social science questions, but scalable analysis requires automation to be practical—and such automation via computer vision methods may uncover important signals not readily accessible by humans (Naik et al., 2017).

Automated explorations of such high-dimensional and unstructured data for social science present some unique challenges that are absent or less important in other empirical fields; these challenges have often been neglected by work in, for example, data mining, even when applied to behavioral data. Social science is generally concerned, like other sciences, with learning about causal relationships, but these are frequently confounded by highly heterogeneous units selecting into different exposures and behaviors based on their expectations (e.g., workers selecting into job training, social media users selecting into exposure to misinformation). Thus, automated generation of insights should aim to (a) surface to researchers empirical regularities that are useful for causal inference while adjusting for confounding and (b) help researchers identify sources of plausibly exogenous (i.e., as-good-as-random) variation in variables of interest.

6.2. Automated Meta-analysis

One area where there is considerable potential for automation to generate new insights without requiring major advances in technology is automated synthesis of large amounts of data via meta- or mega-analysis. Examples of such approaches are common in the biomedical sciences: in genetics, researchers have automatically conducted genome-wide association studies (GWAS) of thousands of phenotypes (McInnes et al., 2018); in neuroimaging, large-scale automated meta-analyses of functional MRI data have been conducted for hundreds of concepts (Yarkoni et al., 2011). Although the social sciences deal with less well-defined entities, there are many domains where even relatively simple automated meta-analysis tools could potentially provide valuable benefits. Limited forms of automated meta-analysis should be almost immediately feasible in virtually any domain where researchers have converged on, and widely adopted, a common ontology.

To illustrate, consider the question of whether and how neuroticism—a personality trait defined by the tendency toward frequent and intense negative emotion—influences physical health (Lahey, 2009; Smith, 2006). There are literally hundreds of publicly accessible data sets containing at least one neuroticism-related measure and at least one measured health outcome. Given controlled vocabularies stipulating what variables can be treated as proxies of neuroticism and health, respectively, it would be relatively straightforward to write meta-analytic algorithms that automatically analyze all such data sets and produce a variety of meta-analytic

estimates (e.g., automatically conducting separate analyses for self-reported versus objectively measured health outcomes; for different sexes and populations; etc.). While such estimates would undoubtedly be of considerably lower quality than comparable manual analyses, they would be effortlessly scalable, and could provide researchers with powerful new tools for large-scale exploration and hypothesis generation.

6.3. Reasoning Systems

At present, truly creative scientific inference remains the sole province of human beings. However, there are reasons to believe that machine intelligence may, in limited domains, begin to complement human reasoning systems in the coming decades. In mathematical and computational fields, where theories deal with more precisely defined entities and much of inference is deductive in nature, there are many examples of automated systems solving nontrivial problems—most commonly under the guise of automated theorem proving (Loveland, 2016). However, isolated examples of automated systems inductively discovering new insights with little or no human intervention can be found in the biological sciences too (e.g., R. D. King et al., 2009).

Broadly speaking, we envision two general ways in which automated scientific reasoning technologies could be introduced into the social sciences. The first resembles existing approaches in other fields, in that it relies largely on the application of explicit rule-based inference systems (sometimes called ‘inference engines’) to codified, formal representations of scientific entities. In the social science context, an autonomous reasoning system of this sort might proceed by (i) detecting novel regularities in available data, (ii) encoding these regularities as formal hypotheses, (iii) generating quantitative predictions and identifying suitable additional data sets to test them in, and (iv) carrying out those tests and drawing appropriate, statistically grounded conclusions. As a speculative example, consider the neuroticism–health relationship described in the previous section. Given a sufficiently comprehensive formal representation of the personality and health domains, once an automated system establishes that neuroticism is correlated with poorer health, it could potentially go on to generate and test specific causal models relating the two. For example, upon recognizing that NEUROTICISM is associated with greater STRESS, that STRESS triggers COPING STRATEGIES, and that some COPING STRATEGIES like SMOKING, DRUG USE, or COMPULSIVE EATING are associated with poor HEALTH, the system might formulate a set of causal graphical models that can be explicitly tested and compared using available data sets.

The second route to automated inference eschews explicit inference rules and instead takes inspiration from numerous machine learning applications that have successfully solved important real-world problems using deep neural network models. Such models are typically trained to maximize performance on some well-defined, predictive criterion; however, once acceptable performance is achieved, the focus then often shifts to generating low-dimensional interpretations of the learned representations, providing human experts with potentially valuable insights (Montavon et al., 2018; Olah et al., 2018). In this vein, we could imagine an extremely complex ‘black-box’ neural network that can nontrivially predict future economic performance from current economic and political indicators. Even if the internal dynamics of the model were too complex for

humans to truly understand, probing the system in the right way might reveal simpler approximations that yield novel scientific hypotheses and help guide policy decisions.

6.4. Outlook

While limited forms of automated insight generation are probably already feasible in some domains of social science, most of the applications discussed here depend on technological and standardization advances that may take decades to realize and widely adopt. Rather than attempting to directly develop reasoning systems that display something akin to scientific creativity, we think such applications may be most useful as aspirational long-term objectives that can help constrain and guide short-term efforts of the kinds discussed in the previous sections. Viewing automated insight generation as a long-term aspirational goal should also hopefully provide the motivation and justification for researchers to carefully consider a number of important pragmatic and ethical considerations that the realization of such technologies might introduce—a topic we turn to next.

7. Practical and Ethical Considerations

Our review has focused predominantly on technical challenges and opportunities associated with efforts to automate social science research workflows. But as we have already observed above, the introduction and widespread use of automation in the social sciences will undoubtedly also bring into focus a range of practical and ethical considerations. We focus on two in particular here: first, the importance of adoption-centric thinking, and second, the need for continual calibration and human oversight.

7.1. Practical Considerations: Adoption and Training

We have already highlighted a number of practical considerations relevant to the technologies and methods we discuss. Here we discuss two in more detail: adoption and training. The movie *Field of Dreams* famously popularized the aphorism that “if you build it, they will come.” Unfortunately, in the world of scientific tool development, they (usually) do not come just because you build it. Open-source software repositories and methods-oriented journals are littered with the abandoned husks of ‘state-of-the-art’ tools that once objectively performed well on some seemingly important metric, but were never widely adopted by the community. Addressing the proliferation of scientific tools in the face of limited user attention is a critical problem that cuts across all of the challenges we discuss above. Unfortunately, it’s also one that funders and researchers alike tend to overlook—arguably because many of the tasks involved, however important, can seem peripheral to the core scientific enterprise.

To address this, scientific tool developers should take user experience seriously. Most scientists are unlikely to learn a new programming language just to use a new tool, so it’s important for tool developers to lower barriers to access whenever possible—for example, by wrapping one’s robust but hard-to-use software library in intuitive graphical interfaces. Relatedly, it’s difficult to exaggerate how important good documentation is for

tool adoption. Ideally, a complex new tool or standard should be accompanied not only by a technical reference, but also by interactive tutorials, worked examples, and user guides that target potential users at all levels of expertise.

Tools and platforms should also be designed in a way that provides immediate benefits to new users (Roure & Goble, 2009). Many a state-of-the-art tool has floundered because its creators were thinking primarily about its long-term benefits, once users have conquered a potentially steep learning curve and invested considerable energy transitioning to a new system. Appeals to altruism or long-term gains are much less compelling to most people than concrete demonstrations of short-term benefits. For example, it's probably a mistake to try to pitch a new ontology to researchers by observing that if everyone were to use this ontology, things would be great. Better to release the ontology alongside one or two easy-to-use software packages or web applications that provide immediate benefits to users if they take the small step of annotating their data—for example, by automatically conducting basic quality-control checks, identifying related public data sets, and so on.

Another consideration concerns the challenge of training computationally adept social scientists of the future. A common retort to calls for additional training in the social sciences is ‘What would you take away to make room?’ We should not underplay the value of deep domain knowledge, lest we simply add computational scientists to the butt of the joke about the naive physicist entering a new discipline (<https://xkcd.com/793/>). Social science is increasingly an interdisciplinary endeavor, and such interdisciplinarity will require not merely more specialization, but also more integration between experts with complementary domain knowledge. Institutional inertia is legendary, so reengineering training programs to produce not just skilled social scientists but functional teams of social scientists will be an important consideration for the universities of the future.

7.2. Ethical Considerations: Oversight and Originality

Automation does not imply independence from human oversight. Most automated tools that address social scientists' needs will likely require continual evaluation and calibration in order to maintain high performance and minimize undesirable consequences. The fact that a technology works well in one context does not guarantee successful generalization to other contexts; for example, a statistical validation tool that operates precisely as expected in one social science domain may fail catastrophically if uncritically applied to another domain where one or more of its assumptions are routinely violated. The probability of generalization failures may further increase in cases where a technology relies heavily on black-box components—as in recent high-profile examples of machine learning algorithms producing decisions that bias against certain groups (Corbett-Davies et al., 2017; Johndrow & Lum, 2017). Consequently, it is imperative that researchers who are working to publicly deploy new automated technologies also simultaneously develop rules and guidelines for continued quality assurance—including specific operating requirements and concrete contingency plans for addressing potential failures (Brundage et al., 2018).

At the same time, it is important to remember that human analysts are also prone to error and bias, and that the actual causes of human-issued decisions also often defy simple explanation (i.e., a good deal of human behavior is also arguably a product of black-box reasoning). Automated technologies need not be perfect to merit widespread deployment; they need only have more favorable cost-benefit profiles than corresponding manual approaches. Of course, this is only true in the case of well-defined problems—the kind at which computers presently excel. Automated tools are not particularly good at defining the problems in the first place, and we must not undervalue the importance of humans for their ability to find those problems. This applies both to the creative nature of hypothesis formation and to the societal impact of automation whenever research has technological or policy implications. The latter often requires an on-the-ground understanding of how people interact with their physical and social environments. For example, when automated systems used by police to allocate patrols are calibrated using existing arrest rates, the results often reinforce racial biases due to the inclusion of petty crimes such as loitering (O’Neil, 2017). The price of automation is eternal vigilance.

8. Conclusion

Automation plays a central and rapidly growing role in our daily lives; it has influenced virtually every domain of human activity, and science is no exception. But the impacts have been uneven, with the social sciences presently lagging far behind the natural and biomedical sciences in the use of automation to solve practical and theoretical challenges. While the entities studied by the social sciences may be more difficult to operationalize and formalize than those found in many other sciences, we argue that social scientists could nevertheless be harnessing automation to a far greater extent. Our review highlights a number of important challenges in social science that increased automation could help mitigate or even solve—including some that appear relatively tractable in the near term. Greater investment in automation would be an effective and extremely cost-efficient way to boost the reliability, validity, and utility of a good deal of social science research.

Acknowledgments

This paper emerged from discussions at a 2-day DARPA-sponsored workshop held at the Center for Open Science (COS) in September 2018. The authors are grateful to John Myles White for valuable discussion and feedback.

Disclosure Statement

Work was partially supported by NIH award R01MH109682 to TY, and by awards to JIL from the Overdeck Family Foundation, Eric and Wendy Schmidt by recommendation of the Schmidt Futures program, and the Alfred P. Sloan Foundation.

References

- Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharsk, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., . . . Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, 4(1), 4–6. <https://doi.org/10.1038/s41562-019-0772-6>
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltmann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Antenucci, D., Cafarella, M. J., & Levenstein, M. C. (2013). Ringtail: a generalized nowcasting system. *Proceedings of the VLDB Endowment*, 6(12), 1358–1361. <https://doi.org/10.14778/2536274.2536315>
- Baker, M. (2016). Stat-checking software stirs up psychology. *Nature*, 540(7631), 151–152. <https://doi.org/10.1038/540151a>
- Bambauer, J., Muralidhar, K., & Sarathy, R. (2014). Fool’s gold: An illustrated critique of differential privacy. *Vanderbilt Journal of Entertainment and Technology Law*, 16(4), 701–728. <https://scholarship.law.vanderbilt.edu/cgi/viewcontent.cgi?article=1207&context=jetlaw>
- Bar-Ilan, J., & Halevi, G. (2018). Temporal characteristics of retracted articles. *Scientometrics*, 116(3), 1771–1783. <https://doi.org/10.1007/s11192-018-2802-y>
- Bender, S., Hirsch, C., Kirchner, R., Bover, O., Ortega, M., L. D’Alessio, Dias, T., Guimarães, P., Lacroix, R., Lyon, M., et al. (2016). *Inexda—the granular data network* (Technical report). IFC Working Papers.
- Berman, F., & Crosas, M. (2020). The research data alliance: Benefits and challenges of building a community organization. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.5e126552>
- Boettiger, C. (2015). An introduction to docker for reproducible research. *Operating Systems Review*, 49(1), 71–79. <https://doi.org/10.1145/2723872.2723882>
- Brainard, J. (2020, May 13). Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? *Science*. <https://doi.org/10.1126/science.abc7839>
- Brown, N. J. L., & Heathers, J. A. J. (2016). The GRIM test. *Social Psychological and Personality Science*, 8(4), 363–369. <https://doi.org/10.1177/1948550616673876>

- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Chess, B., & McGraw, G. (2004). Static analysis for security. *IEEE Security Privacy*, 2(6), 76–79. <https://doi.org/10.1109/MSP.2004.111>
- Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920–980. <https://doi.org/10.1257/jel.20171350>
- Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from largescale biology. *Science*, 300(5617), 286–290. <https://doi.org/10.1126/science.1084564>
- Commission on Evidence-Based Policymaking. (2017). *CEP final report: The promise of evidence-based policymaking* (Technical report).
- Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797–806). <https://doi.org/10.1145/3097983.3098095>
- Dafoe, A. (2014). Science deserves better: The imperative to share complete replication files. *PS Political Science & Politics*, 47(1), 60–66. <https://doi.org/10.1017/s104909651300173x>
- Dankar, F. K., & El Emam, K. (2013). Practicing differential privacy in health care: A review. *Transactions on Data Privacy*, 6(1), 35–67. <http://www.tdp.cat/issues11/tdp.a129a13.pdf>
- Duşa, A., Nelle, D., Stock, G., & Wagner, G. G. (2014). *Facing the future: European research infrastructures for the humanities and social sciences*. Allea. https://www.allea.org/wp-content/uploads/2015/09/2014_06_04-FACING_THE_FUTURE.pdf
- Dwork, C. (2008). Differential privacy: A survey of results. In M. Agrawal, D. Du, Z. Duan, & A. Li (Eds.), *Lecture Notes in Computer Science: Vol. 4978. Theory and applications of models of computation: TAMC 2008* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210), Article 1243089. <https://doi.org/10.1126/science.1243089>
- Elias, P. (2018). The UK administrative data research network: Its genesis, progress, and future. *The ANNALS of the American Academy of Political and Social Science*, 675(1), 184–201.

<https://doi.org/10.1177/0002716217741279>

Epskamp, S., & Nuijten, M. B. (2014). Statcheck: Extract statistics from articles and recompute p values (R package version 1.0. 0.).

Gerber, A., & Malhotra, N. (2008). Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science*, 3(3), 313–326.

<https://doi.org/10.1561/100.00008024>

Gewin, V. (2014). Retractions: A clean slate. *Nature*, 507(7492), 389–391. <https://doi.org/10.1038/nj7492-389a>

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676. <https://doi.org/10.1016/j.jmoneco.2008.05.010>

Gorgolewski, K., Esteban, O., Schaefer, G., Wandell, B., & Poldrack, R. (2017). OpenNeuro—A free online platform for sharing and analysis of neuroimaging data. In *Organization for human brain mapping* (no. 2, p. 1677). <https://doi.org/10.7490/F1000RESEARCH.1114354.1>

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., Churchill, N. W., Cohen, A. L., Craddock, R. C., Devenyi, G. A., Eklund, A., Esteban, O., Flandin, G., Ghosh, S. S., Guntupalli, J. S., Jenkinson, M., Keshavan, A., Kiar, G., Liem, F., ... Poldrack, R. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Computational Biology*, 13(3), Article e1005209. <https://doi.org/10.1371/journal.pcbi.1005209>

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, Article 160044. <https://doi.org/10.1038/sdata.2016.44>

Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *BMJ*, 339, Article b2680. <https://doi.org/10.1136/bmj.b2680>

Habibi, M., Weber, L., Neves, M., Wiegandt, D. L., & Leser, U. (2017). Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics*, 33(14), i37–i48. <https://doi.org/10.1093/bioinformatics/btx228>

Hales, B. M., & Pronovost, P. J. (2006). The checklist—A tool for error management and performance improvement. *Journal of Critical Care*, 21(3), 231–235. <https://doi.org/10.1016/j.jcrc.2006.06.002>

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A.-H. S., Dellinger, E. P., Herbosa, T., Joseph, S., Kibatala, P. L., Lapitan, M. C. M., Merry, A. F., Moorthy, K., Reznick, R. K., Taylor, B., Gawande, A. A., & Safe Surgery Saves Lives Study Group. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine*, 360(5), 491–499.

<https://doi.org/10.1056/NEJMsa0810119>

Heathers, J. A., Anaya, J., van der Zee, T., & Brown, N. J. L. (2018). *Recovering data from summary statistics: Sample Parameter Reconstruction via Iterative TEchniques (SPRITE)* (Technical report e26968v1). PeerJ Preprints. <https://doi.org/10.7287/peerj.preprints.26968v1>

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70–85. <https://doi.org/10.1177/2515245917751886>

Inter-university Consortium for Political and Social Research. (2009). *Guide to social science data preparation and archiving: Best practice throughout the life cycle*. Inter-University Consortium for Political & Social Research.

Ioannidis, J. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19(5), 640–648.

<https://doi.org/10.1097/ede.0b013e31818131e7>

Ioannidis, J. P. A. (2005). Why most published research findings are false (W. Jantsch & F. Schaffler, Eds.). *PLOS Medicine*, 2(8), Article e124. <https://doi.org/10.1371/journal.pmed.0020124>

Jarmin, R. S., & O'Hara, A. B. (2016). Big data and the transformation of public policy analysis. *Journal of Policy Analysis and Management*, 35(3), 715–721. <https://doi.org/10.1002/pam.21925>

Johndrow, J. E., & Lum, K. (2017). An algorithm for removing sensitive information: Application to race-independent recidivism prediction. *The Annals of Applied Statistics*, 13(1), 189–220. <https://doi.org/10.1214/18-aos1201>

Kim, S., Fiorini, N., Wilbur, W. J., & Lu, Z. (2017). Bridging the gap: Incorporating a semantic similarity measure for effectively mapping PubMed queries to documents. *Journal of Biomedical Informatics*, 75, 122–127. <https://doi.org/10.1016/j.jbi.2017.09.014>

King, G. (2014). Restructuring the social sciences: Reflections from Harvard's Institute for Quantitative Social Science. *PS: Political Science & Politics*, 47(1), 165–172. <https://doi.org/10.1017/s1049096513001534>

King, G., Honaker, J., Joseph, A., & Scheve, K. (1998). *List-wise deletion is evil: What to do about missing data in political science*. Annual Meeting of the American Political Science Association, Boston.

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.199.5654&rep=rep1&type=pdf>

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., & Clare, A. (2009). The automation of science. *Science*, 324(5923), 85–89. <https://doi.org/10.1126/science.1165620>

Loizides, F., & Schmidt, B. (Eds.). (2016). *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*. IOS Press. https://goedoc.uni-goettingen.de/bitstream/handle/1/13343/Loizides_Schmidt.pdf?sequence=1&isAllowed=y

Korpela, K. M. (2010). How long does it take for the scientific literature to purge itself of fraudulent material?: The Breuning case revisited. *Current Medical Research and Opinion*, 26(4), 843–847. <https://doi.org/10.1185/03007991003603804>

Kreuter, F., Ghani, R., & Lane, J. (2019). Change through data: A data analytics training program for government employees. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.ed353ae3>

Kriegeskorte, N., Walther, A., & Deca, D. (2012). An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. *Frontiers in Computational Neuroscience*, 6, 94. <https://doi.org/10.3389/fncom.2012.00094>

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015). From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning—Volume 37* (pp. 957–966). <http://proceedings.mlr.press/v37/kusnerb15.pdf>

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64(4), 241–256. <https://doi.org/10.1037/a0015309>

Lane, J. (2016). Big data for public policy: The quadruple helix. *Journal of Policy Analysis and Management*, 35(3), 708–715. <https://doi.org/10.1002/pam.21921>

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062. <https://doi.org/10.1126/science.aaz8170>

LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org : Grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, 8(4), 424–432. <https://doi.org/10.1177/1745691613491437>

Lee, J., & Clifton, C. (2011). How much is enough? Choosing ϵ for differential privacy. In X. Lai, J. Zhou, & H. Li (Eds.), *Lecture notes in computer science: Vol. 7001. Information security. ISC 2011* (pp. 325–340). Springer. https://doi.org/10.1007/978-3-642-248610_22

Levenstein, M. C., & Lyle, J. A. (2018). Data: Sharing is caring. *Advances in Methods and Practices in Psychological Science*, 1(1), 95–103. <https://doi.org/10.1177/2515245918758319>

Levenstein, M. C., Tyler, A. R. B., & Bleckman, J. D. (2018.) *The researcher passport: Improving data access and confidentiality protection*. Unpublished working paper: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/143808/ICPSR_ResearcherCredentialingWhitePaper_May2018.pdf?sequence=1

Liu, T. X., Yang, J., Adamic, L. A., & Chen, Y. (2014). Crowdsourcing with All-Pay auctions: A field experiment on Taskcn. *Management Science*, 60(8), 2020–2037. <https://doi.org/10.1287/mnsc.2013.1845>

Loveland, D. W. (2016). *Automated theorem proving: A logical basis*. Elsevier.

McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J. E., Aguirre, M., & Rivas, M. (2018). Global Biobank Engine: Enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics*, 35(14), 2495–2497. <https://doi.org/10.1093/bioinformatics/bty999>

Meng, X.-L. (2020). Covid-19: A massive stress test with many unexpected opportunities (for data science). *Harvard Data Science Review*, (Special Issue 1). <https://doi.org/10.1162/99608f92.1b77b932>

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. <https://doi.org/10.48550/arXiv.1301.3781>

Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>

Moran, E. F., Hofferth, S. L., Eckel, C. C., Hamilton, D., Entwisle, B., Aber, J. L., Brady, H. E., Conley, D., Cutter, S. L., Hubacek, K., & Scholz, J. T. (2014). Opinion: Building a 21st-century infrastructure for the social sciences. *Proceedings of the National Academy of Sciences*, 111(45), 15855–15856. <https://doi.org/10.1073/pnas.1416561111>

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), Article 0021. <https://doi.org/10.1038/s41562-016-0021>

National Research Council, Division of Behavioral and Social Sciences and Education, Committee on Population, Committee on National Statistics, Board on Behavioral, Cognitive, and Sensory Sciences, & Committee on Revisions to the Common Rule for the Protection of Human Subjects in Research in the Behavioral and Social Sciences. (2014). *Proposed revisions to the common rule for the protection of human subjects in the behavioral and social sciences*. National Academies Press.

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P., Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, 20(3), 299–303.

<https://doi.org/10.1038/nn.4500>

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin, F. S., 3rd, Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., ... de Zeeuw, A. (2016). Social norms as solutions. *Science*, 354(6308), 42–43.

<https://doi.org/10.1126/science.aaf8317>

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The building blocks of interpretability. *Distill*, 3(3), Article e10. <https://doi.org/10.23915/distill.00010>

O’Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. <https://doi.org/10.1126/science.aac4716>

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press. <https://doi.org/10.1017/cbo9781316423936>

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annual Review of Psychology*, 67, 587–612. <https://doi.org/10.1146/annurev-psych-122414-033729>

Poteete, A. R., Janssen, M. A., & Ostrom, E. (2010). *Working together: Collective action, the commons, and multiple methods in practice*. Princeton University Press.

Roure, D. D., & Goble, C. (2009). Software design for empowering scientists. *IEEE Software*, 26(1), 88–95. <https://doi.org/10.1109/ms.2009.22>

Ruggles, S., Fitch, C., Magnuson, D., & Schroeder, J. (2019). Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109, 403–408. <https://doi.org/10.1257/pandp.20191107>

Sala-i-Martin, X. X. (1997). I just ran two million regressions. *American Economic Review*, 87(2), 178–183. <https://doi.org/10.2139/ssrn.31213>

Schuemie, M. J., Ryan, P. B., DuMouchel, W., Suchard, M. A., & Madigan, D. (2014). Interpreting observational studies: Why empirical calibration is needed to correct p-values. *Statistics in Medicine*, 33(2),

209–218. <https://doi.org/10.1002/sim.5925>

Shen, H. (2014). Interactive notebooks: Sharing the code. *Nature*, 515(7525), 151–152. <https://doi.org/10.1038/515151a>

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. <https://doi.org/10.1037/a0033242>

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), Article 160384. <https://doi.org/10.1098/rsos.160384>

Smith, T. W. (2006). Personality as risk and resilience in physical health. *Current Directions in Psychological Science*, 15(5), 227–231. <https://doi.org/10.1111/j.1467-8721.2006.00441.x>

Steenen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives in Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>

Tamer, E. (2019). The ET interview: Professor Charles Manski. *Econometric Theory*, 35(2), 233–204. <https://doi.org/10.1017/S0266466618000075>

Treloar, A. (2014). The research data alliance: Globally co-ordinated action against barriers to data publishing and sharing. *Learned Publishing*, 27(5), S9–S13. <https://doi.org/10.1087/20140503>

Vardigan, M., Heus, P., & Thomas, W. (2008). Data documentation initiative: Toward a standard for the social sciences. *International Journal of Digital Curation*, 3(1). <https://doi.org/10.2218/ijdc.v3i1.45>

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*, 8(7), Article e67332. <https://doi.org/10.1371/journal.pone.0067332>

Weinberg, D., Abowd, J. M., Belli, R. F., Cressie, N., Folch, D. C., Holan, S. H., Levenstein, M. C., Olson, K. M., Reiter, J. P., Shapiro, M. D., Smyth, J., Soh, L.-K., Spencer, S., Vilhuber, L., & Winkle, C. (2017). Effects of a government-academic partnership: Has the NSF-Census Bureau research network helped secure the future of the federal statistical system? *Journal of Survey Statistics and Methodology*, 7(4), 589–619. <https://doi.org/10.1093/jssam/smy023>

Yarkoni, T. (2012). Designing next-generation platforms for evaluating scientific output: What scientists can learn from the social web. *Frontiers in Computational Neuroscience*, 6, 72. <https://doi.org/10.3389/fncom.2012.00072>

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665–670.

<https://doi.org/10.1038/nmeth.1635>

Ye, X., Shen, H., Ma, X., Bunescu, R., & Liu, C. (2016). From word embeddings to document similarities for improved information retrieval in software engineering. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 404–415). <https://doi.org/10.1145/2884781.2884862>

York, D. G., Adelman, J., Anderson Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S. Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., . . . Yasuda, N. (2000). The Sloan Digital Sky survey: Technical summary. *The Astronomical Journal*, 120(3), Article 1579. <https://doi.org/10.1086/301513>

Young, A. (2015). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2), 557–598.

<https://doi.org/10.1093/qje/qjy029>

Zhang, Y., Rahman, M. M., Braylan, A., Dang, B., Chang, H. L., Kim, H., McNamara, Q., Angert, A., Banner, E., Khetan, V., & McDonnell, T. (2016). Neural information retrieval: A literature review. *arXiv*.

<https://doi.org/10.48550/arXiv.1611.06792>

Zheng, J., Williams, L., Nagappan, N., Snipes, W., Hudepohl, J. P., & Vouk, M. A. (2006). On the value of static analysis for fault detection in software. *IEEE Transactions on Software Engineering*, 32(4), 240–253.

<https://doi.org/10.1109/tse.2006.38>

©2021 Tal Yarkoni, Dean Eckles, James A. J. Heathers, Margaret C. Levenstein, Paul E. Smaldino, and Julia Lane. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/> ↵