Algorithms, Big Data, and Inequality:
Challenging Data-Driven Discrimination by Examining the Design and Use of Algorithmic Systems

ISS Collaborative Project Proposal
2018-2021

Team Members:
Martin T. Wells (Team Leader)
Professor, Departments of Social Statistics and Statistical Science (mtw1@cornell.edu)

Ifeoma Ajunwa
Assistant Professor, Department of Organizational Behavior, ILR School (iajunwa@cornell.edu)

Solon Barocas
Assistant Professor, Department of Information Science (sbarocas@cornell.edu)

Brooke Erin Duffy
Assistant Professor, Department of Communication (bduffy@cornell.edu)

Malte Ziewitz
Assistant Professor, Department of Science & Technology Studies (mcz35@cornell.edu)

Abstract

The proliferation of algorithms is reconfiguring our socio-economic systems in profound and complex ways. This collaborative project brings together scholars from across the social sciences to critically examine the design, understanding, and use of algorithmic systems. The researchers share a concern with how the concomitant rise of big data, machine learning, and digital surveillance has the potential to exacerbate social inequalities among vulnerable communities. The proposed program of research will therefore develop a more holistic understanding of algorithmic bias across disciplinary boundaries and within such empirical domains as work/employment, finances/creditworthiness, health, and social media. Through such research, we aim to understand—and ultimately challenge—the kinds of data-driven inequality and discrimination that are defining social life in the algorithmic age.
Algorithms, Big Data, and Inequality: Challenging Data-Driven Discrimination by Examining the Design and Use of Algorithmic Systems

Background and Project Significance

The emergence of computational techniques like big data analytics and machine learning represents one of the most important social, political-economic, and technological developments of the last two decades. Search engine rankings can make or break a business, reputation, or career (Gandini, 2016), or contribute to racial inequality (Noble, 2018); scoring systems determine our creditworthiness for banks, lenders, and insurance brokers (Robinson and Yu, 2014); and algorithmic hiring platforms parse résumés to calculate who should and, conversely, who shouldn’t be seen by human managers (Smith and Anderson, 2017). In short, algorithmic systems have come to play a crucial role in ordering our lives, even being deployed as a tool of law and order (Eubanks, 2018). Of public and scholarly concern is that these systems may contribute to—and even exacerbate—existing social inequalities and power differentials (Ajunwa, 2017; Barocas & Selbst, 2016; O’Neil, 2016).

Fueled by information garnered from health records (Ajunwa, 2016), advertising and retail activities (Turow, 2012), social media practices (Duffy, 2017), and public records, contemporary institutions have a treasure trove of data. This data often is collected with little regard to privacy and security (Ajunwa, 2014) and potentially discriminatory or unethical uses (Ajunwa, 2017; Barocas and Selbst 2016). Meanwhile, both algorithms and the data feeding them remain ‘black-boxed’—a mix of technical complexity and corporate secrecy that leaves the public vulnerable to unjust and unfair outcomes with little recourse (Kroll et al., 2016; Pasquale, 2015). While the growing academic field of critical data and algorithm studies has begun to examine the design and use of algorithmic systems (Sandvig et al. 2014; Ziewitz 2016), the general public remains largely unaware of the extent of algorithmic influence on social and
economic opportunities. Studies from the Pew Research Center show that, for example, “fewer than half of Americans are familiar with the concept of computer programs that can review job applications without any human involvement” (Smith and Anderson, 2017).

In a recent *NY Times* opinion essay, *Weapons of Math Destruction* author Cathy O’Neil (2017) issued a call to arms for the funding of academic research on ethical issues in algorithmic decision-making. Major media platforms are also giving credence to the idea that algorithmic decision-making risk perpetuating economic inequality (Goodman, 2018). Against this backdrop, our ISS Collaborative Project team will draw upon a range of methods—including in-depth interviews, participant-observation, statistical analysis, policy analysis, and surveys—to critically examine the *design* and *use* of algorithmic systems. We take a sociotechnical systems and organizational studies approach to the relationships between technologies, organizations, and human actors and thus propose to examine algorithmic systems at various moments in the process of *design* and *use*. Owing to the intellectual diversity of our team, our particular programs traverse various institutions and contexts and our research arenas include: hiring and workplace monitoring, social media platforms, user scoring, healthcare, manufacturing, and farming. Together, they will help us understand how design and use interact to create algorithmically vulnerable communities—that is, social groups at greatest risk of unfair or objectionable judgments and least resourced to understand and challenge them.

**Research Plan and Design**

Our proposed program of research examines how the design, use, and public understanding of algorithmic systems contributes to existing inequalities or creates new socio-economic inequalities in specific empirical domains. While our extant scholarship (e.g., Ajunwa 2016; Ajunwa, 2017; Ajunwa et. al, 2017; Barocas et al., 2013, Barocas and Selbst 2016; Duffy
2017; Ziewitz 2016) has begun to document the negative social and economic consequences of algorithmic systems, substantial work remains. Team members will therefore collaborate to advance scholarly understanding of the role of algorithmic systems in contributing to inequality, calling particular attention to the complex relationship between design and use—how these systems are made to work in practice, how practice differs from imagined use, and how the experience of these systems can and should inform design.

Each of the four proposed projects examines a distinctive layer of inequality in algorithmic systems in order to contribute to a comprehensive social scientific understanding of data-driven discrimination across social, organizational, and institutional contexts. The first project examines how bias may get encoded into the design process in the employment market, which provides a useful segue into the second project’s audit of hiring algorithms. The third project shifts the focus to how various publics understand—and respond to–algorithmic systems. Finally, the fourth project provides a comparative dimension to the other components by examining algorithmic engagement in the global south.

A. Discourses of Design and Use

In the wake of profound transformations in post-industrial labor markets and economies, the features and conditions of work are being radically reconfigured. Yet, traditional inequalities—including gender (Blau and Kahn, 2006), race, and class origin (Laurison and Friedman, 2016)—persist in the so-called “new economy” in the form of wage gaps, horizontal segregation/occupational clustering (Charles and Grusky, 2004), and more insidious forms of discrimination covertly expressed via algorithmic decision-making (O’Neil, 2016). Technologies of the information age promise (superficially, at least) to challenge these inequalities. Indeed, amidst a hyper-saturated employment market, time-starved hiring managers are increasingly
relying on data-driven technologies to target and screen candidates: “hiring algorithms,” in particular, are pitched to companies as a way to increase efficiency, aggregate information, and above all, reduce the human bias implicit in the job screening process. Yet evidence from various industries—from the low-wage work to white collar work—suggests that these egalitarian claims may be inflated. That is, these data-driven programs may enable more subtle forms of discrimination that may exacerbate existing social hierarchies (Barocas and Selbst 2016; Ajunwa et al, 2017; Peerson, 2016). To take one recent example, research has demonstrated that models developed using natural language processing (to potentially inform the automatic assessment of job applicants) may learn gender and racial stereotypes that echo those held by humans (Caliskan-Islam et. al, 2017).

Accordingly, this research will perform a critical survey of the ecosystem of algorithmic work technologies that already exist in the labor market. This collaborative project will also examine how various stakeholders understand the role of hiring algorithms and productivity applications in creating a more (in)equitable labor market. This research will build upon Ajunwa’s qualitative research on work algorithms along with Duffy’s (2017) recently published study of the gendered language of social media job advertisements to explore the intersection of work, technology, and inequality. Specifically, this research will examine the extent to which various types of bias get coded into these algorithmic systems through an analysis of various stakeholders in the hiring process, including technologists who create these programs, hiring managers, and the workers themselves.

The qualitative data from this project will come from two case studies, one investigating hiring by white-shoe firms in the financial industry, and another examining the media and creative industries. While the former is known for its selectivity, the creative industries are
marked by persistent instability. Crucially, the so-called “politics of insecurity” (Beck, 2000) affect categories of workers unevenly: low-wage and low-status workers are especially disadvantaged by the forces of worker individualization (e.g., Smith, 2016; Vosko, 2010), but we posit that with the advent of algorithmic hiring tools (and its requirement for discrete variables derived from big data), higher wage workers will come to experience what Ajunwa terms, “the quantification of the worker.”

To allow for comparisons, the two case studies will revolve around the same research questions, including: Who do designers imagine as the user of their tech products? What stories do they tell about the role of their products in economy? To what extent do they acknowledge or make efforts to reduce algorithmic biases? To what extent do managers express concerns about how their use of these algorithmic working tools might reduce or enable bias and inequality? Relatedly, we will develop an understanding from the perspective of managers by assessing: How do managers use these tech products? Do they use them as advertised or is their practical drift in regards to actual use? What are the common hopes/complaints/anxieties that managers have about their use of these algorithmic products? What feedback/requests do managers have for designers? To what extent do managers express concerns about how their use of these algorithmic working tools might reduce or enable bias and inequality?

B. Algorithmic Audit

Nearly all Global 500 companies now use Automated Hiring Platforms (AHPs) (Baber, 2006). A survey of application policies revealed that the top 20 private employers in the US, as ranked in the Fortune 500, all require job applications to be submitted online (Greene and Ajunwa, forthcoming). Automated hiring platforms thus represent inescapable checkpoints for large swathes of the labor market, particularly those seeking hourly retail positions with retail
corporations. There is also growing evidence for the popularity of hiring algorithms for white-collar workers among business firms (Gee, 2017). Yet, there is some research to suggest that algorithms may not be neutral and may instead evince the racial, gender, and other biases of their human creators (Mann & O’Neil, 2016).

While hiring algorithms might seem like a way to overcome the prejudices, implicit biases, and faulty heuristics that plague human decision-making, rules learned from historical data can easily inherit the prejudices of prior decision makers, reflect the widespread biases that persist in society, or discover useful regularities that are really just pre-existing patterns of exclusion and inequality (Ajunwa, 2018 forthcoming; Barocas and Selbst 2016). Our proposal focuses on two particular issues: (i) diagnosing and auditing bias in black-box algorithms and (ii) algorithmic approaches to fairness aware learning. This proposal builds on work by team members, Ifeoma Ajunwa and Martin Wells, who have received a theme project grant from the ILR School as seed money for this sub-project.

A simple approach to detect bias in machine-based scoring models could be to reverse engineer them. However, this will be foiled by the lack of access to all features (known as predictor variable in regression modeling) and the same data sample used to create the model, or a means of determining how close the result is to the true, unknown model. On the other hand, one could study the actual outcome, testing for disparate impact through methods such as training a model to predict the outcome, removing, permuting, or obscuring a protected feature (Feldman et al., 2015; Adler et al 2016), and then retraining the model to see if it changes. One challenge with this approach is that advance knowledge of which feature to act on is needed, and there may be biases in the data that are not a priori known. Chouldechova and G’Sell (2017) proposed an approach to identify subgroups where two classifiers differ in terms of fairness,
recursively partitioning covariate space by testing for homogeneity of a fairness metric of interest (e.g. false positive rate) between the two classifiers.

As a first project, we propose to use a parsimonious modeling approach for bias detection that examines both the prediction score as well as the actual outcome, leveraging the difference between them to detect potential bias (Tan et al. 2017). This approach involves training two separate models: a transparent low dimensional model to mimic a black-box prediction score and another transparent model to predict the actual outcome that the prediction score was intended to predict. We then use a standard statistical procedure to assess systematic differences in the predictive scoring model compared to the actual outcome. This approach can be applied to generic black-box models where one does not have knowledge of model features used to make predictions, and does not require the models to use the same input features. It is of particular interest when systematic differences occur on protected features. We also plan to establish some inferential and computational performance guarantees for this approach.

We also propose to develop regularization approaches for a large number of features that are adapted for the fairness constraints as well guard against overfitting. The algorithms developed in Schifano et al. (2010) and Bar et al. (2017) are sufficiently flexible to deal with the shortcoming of previous approaches and scale to large number of features. As an application, we will develop new fairness-aware predictive scoring algorithms that can be applied in many domains in addition to hiring, such as financial services, health care, and criminal justice. In surfacing worrisome outcomes, this project will naturally feed into the previous project, which aims to evaluate whether these outcomes conform to legal requirements and ethical principles.
We envision that this work could have crossovers and potential for further collaboration with members of the Mass Incarceration Collaborative Theme Project, particularly in regards to the use of algorithms in law enforcement and criminal justice decision-making.

C. User Studies and Engagement

While there is growing awareness of the work of algorithms and big data, little is still known about how ordinary users—i.e. those who lack the special expertise of lawyers, policy-makers, and engineers—come to understand and interact with them. To address these issues, this part of the project will expand the focus of analysis by studying algorithms and big data from the perspective of users: how do those who are affected by these systems on a daily basis make sense of and engage with them? What issues of inequality, fairness, and due process arise beyond the technical domain? And what, if anything, can be done to improve the situation? Answering these questions is an important part of understanding inequality in algorithmic systems. While the practices of ordinary users may seem insignificant on their own, they constitute the everyday experience of millions of people and can have powerful consequences in the aggregate.

Three of the team members are in the process of collecting data in this area. Ziewitz has focused on two empirical settings, in which ordinary users play an important but often overlooked role: web search and credit scoring. His work includes 12 months of completed ethnographic fieldwork within the search engine optimization (SEO) industry as well as an ongoing interview-diary study of how people in Upstate New York make sense of and recover from a broken credit score. Duffy is in the midst of conducting in-depth interviews with content creators about how they understand and respond to shifts in platform algorithms. As part of her book in progress, *The Quantified Worker* (Cambridge University Press forthcoming), Ajunwa is interviewing the creators and users of popular work algorithms (specifically, the algorithms
involved in managing hiring, productivity, scheduling, etc.). Working across these fields, the team will explore three areas of inquiry previously developed by Ziewitz:

1. **Folk theories of algorithmic systems:** While ordinary people may not be able to account for scoring algorithms in terms of mathematics and computer science, they nevertheless develop an idea of how these systems work. These ‘folk theories’ may seem insignificant at first, but shape the everyday experiences of millions of users and their capacity to act upon a system. How do ordinary people make sense of systems that appear to be inscrutable?

2. **Strategies of resistance:** In an age of ubiquitous scoring, rating, and ranking, what strategies of resistance and repair, if any, do individuals undertake if they are concerned about their reputations? What are the limitations of these strategies? Studying these practices allows us to tackle ethical concerns in algorithmic systems as they are being negotiated in practice.

3. **Role of intermediaries:** A common phenomenon in algorithmic systems is the development of so-called ‘shadow industries,’ i.e. a new cast of professionals that take advantage of the inscrutability of systems by charging users for helping them optimize their standing vis-à-vis an algorithmic system (Ziewitz 2017). How and under what conditions do people resort to such providers?

Not only will this phase of research shed light on the public understanding of algorithmic systems in various realms of life, including credit scoring, financing, and reputation-building. We will also interrogate the role of user agency in hopes of developing best practices for citizen-consumers in the face of pervasive monitoring and rating practices.
D. International Comparison

The research on algorithms has thus far been dominated by a focus on their use in the North American and European context. Yet, algorithmic processes are impacting the lives of individuals in global south. The rise of algorithmic processes that allow for platform work, for example, has led to what some scholars term the “planetary global market” (Graham et al. 2017), in which algorithms have connected distant labor markets in Africa and Asia with those in the global north. While much has been written about the oversupply of labor on digital platforms as a result of a planetary global market (Graham et al. 2017), there is little research on the experience of workers who serve as draft horses for the algorithmic processes that power the economic activity of the global north. Through this collaborative project, we will begin to explore the impact of algorithms in the labor markets of the global south, notably Asia and Africa.

Mark Graham writes in the New Statesman:

A few months ago, I visited an artificial intelligence training centre in a rural town in Central Africa. Getting there involved a day-long drive from the nearest international airport. It was here, in a place where many people still live in thatched huts and few families possess any of the technological gadgets of the contemporary world, that workers are helping to build some of the world’s most advanced technologies and services. In a large open-plan office with hundreds of desks and computers, workers spend eight hours a day doing highly repetitive work like matching names to photographs of minor celebrities they’ve never heard of, or identifying objects in photos of suburban America in cities that they will never go to. What these tasks have in common with the dozens of other routines performed in the room is that computers cannot yet perform them as effectively as humans. Real people are needed to structure, classify, and tag an enormous amount of unstructured information for companies using machine learning algorithms in products like autonomous vehicles and next-generation search engines.  
(Graham2018)

This part of our project will investigate the phenomenon of “data labeling as the blue collar job of the future” and how such work is outsourced to developing nations with cheaper
labor pools. One of the collaborators, Ifeoma Ajunwa, has a personal interest and business contacts in Africa, in particular, Lagos, Nigeria. The collaborative grant would support exploratory ethnographic research in Nigeria, Kenya, and South Africa that could serve as the basis for applications for further funding. The hope also is that this part of your proposed project will lead to interesting collaborations with the China Cities Collaborative Project, which includes Ajunwa’s ILR School colleague Eli Friedman. Some comparative research questions include: What is the experience of blue-collar algorithmic work in the global south? How do workers navigate tasks that cross language and cultural lines? How does the emergence of such work enable or mitigate larger societal inequalities?

There is, of course, no single story of the impact of technological innovation in the Global South. Alongside research on the role of algorithms in creating distant labor markets of workers, this part of our project would also critically examine the rise of local start-ups powered by algorithms. A Newsweek article provides an eye-opening survey of the African start-up scene noting that “African tech startups increased by a factor of 10 from $41 million in 2012 to $414 million in 2014, and is expected to rise to more than $600 million by 2018.” (Gaffey 2016). Notably, the companies are tapping into the power of algorithms to offer innovative services. Xineoh is profiled as a prominent example of a successful algorithm-powered tech start-up: as it has “generated in excess of $30 million in revenue for clients across the globe” and is regarded as a “pioneer in the AI and machine learning space with their distinctive algorithm bearing similarities to those used by the likes of Amazon.com and Netflix…” (IT News Africa 2017).

These news articles, however, leave several social scientific research questions unanswered. Notably: 1. What are the ideal conditions that lead to the emergence of such start-ups? 2. What are the organizational constraints that stymie the progress of such start-ups? 3.
How do such startups, powered by algorithms, mitigate, exacerbate, or create new social inequalities? These questions demand qualitative research in the form of ethnography and in-depth interviews of start-up entrepreneurs, the workers they recruit, and the investors who fund these startups.

**Timeline**

As all team members have active research agendas related to these questions, our project is shovel-ready. We are well prepared to channel our existing work and working relationships into a cross-disciplinary collaboration that holds the promise of delivering unique insights into the relationship between algorithms and inequality.

To achieve this goal, we will devote **Year 1** to scoping the field, reviewing the literature during regular collaborative meetings, and starting the research collaborations. We will spend this year meeting bi-weekly to develop our shared research agenda and start data collection.

**Year 2,** we will focus on scholarship, teaching, and expanding the visibility of the project. To support this work, we will hire graduate research assistants. Our teaching efforts will build on Professors Ajunwa, Barocas, Duffy, Wells, and Ziewitz’s popular courses touching on algorithms, big data, ethics, and design. To increase the visibility of our project, we will host a special workshop attached to the Fairness, Accountability, Transparency (FAT*) conference, chaired by Barocas, that focuses specifically on empirical social scientific work on algorithms and inequality and solicit contributions for a special journal issue on the topic. We further plan to publicize our collaborative work through a dedicated website with resources and information, editorials in newspapers, and posts on popular academic blogs. By the end of Year 2, we expect to have submitted at least three collaborative research grants through institutions such as the National Institutes of Health, the National Science Foundation, the National Institute of Justice,
the Russell Sage Foundation, and the Department of Labor. The team members have successfully received funding from several of these organizations in the past (see bio sketches attached). We believe that, as an established collaborative group, the chances of success will increase further. These grant applications will help team members identify shared research interests and conduct pilot studies, where appropriate, while providing concrete deadlines.

In **Year 3**, we plan to bring our research projects to the writing stage and to publish research in top peer-reviewed journals. We will also conclude Year 3 with a capstone lecture.

**Outcomes and Final Products**

Our proposed ISS Collaborative Project on Algorithms, Big Data, and Inequality will generate a number of outcomes through scholarly publications, teaching/training, and public engagement. We plan to create an in-depth overview of a neglected important area of research. We therefore aim to place co-authored papers in top peer-reviewed journals in our fields by the end of the project. In particular, a target journal is *American Sociological Review* (ASR), a top sociology journal with wide readership. Other appropriate journals for our disciplines include: *Management Science, Administrative Science Quarterly, Information, Communication & Society, Social Studies of Science, Computer-Supported Cooperative Work, MIS Quarterly*, etc. Through impactful publications in top journals and a dedicated website with resources and information for further research on algorithms, we will ensure that of our collaboration has broad and lasting impact.

In addition to contributing to the scholarly literature, we will address public policy, design, and civil society responses to inequality through white papers and testimonies at the relevant governmental agencies. We plan to engage relevant stakeholders in policy, user, and
academic communities in ongoing conversations with the goal of (re-)shaping the discourse and policy around algorithms.

Our project holds great promise for establishing Cornell as a hub for research excellence in this domain. As a team from five departments and a range of academic disciplines and methodological orientations, we will connect researchers across the social and computational/data sciences at Cornell. We also hope to foster ties between the Ithaca campus and the NYC campuses, including Cornell Tech and Weill Medical College. Finally, we believe that our project would put us on an ideal path to apply for future funding through the National Science Foundation, the Department of Labor, the National Institutes of Health, and the Russell Sage Foundation.

Team Members

Our collaborative team includes Martin Wells (Departments of Social Statistics and Statistical Science), Ifeoma Ajunwa (ILR School and Law School), Solon Barocas (Information Science), Brooke Erin Duffy (Communication), and Malte Ziewitz (Science & Technology Studies). As team leader, Wells will draw on his extensive experience of administering major research grants and managing a range of institutions, including the Cornell Survey Research Institute, the Departments of Biological Statistics and Computational Biology, Social Statistics, and Statistical Science, and various multidisciplinary teams at ILR, Cornell Law School, and Weill Medical College. Professor Wells will further ensure that assistant professors on the team will not be burdened by any logistical and administrative responsibilities while having full input at all stages of the project. We note that our team’s composition of one senior and four junior faculty members is a product of the relatively recent emergence of the field of critical algorithm studies—a development in which the junior members have played a central part. Moreover, two
of the project members (Duffy and Ziewitz) are advanced assistant professors who plan to submit their tenure materials midway through this project.

The team members bring together a unique mix of scholarly expertise, diversity of disciplinary perspectives, and a demonstrated ability to work together. Scholarly expertise is demonstrated in a number of high-profile publications and successful grant applications relevant to this project (see bio sketches attached). For example, Professors Ajunwa and Wells have already begun collaborating as evidenced by an ILR School grant on auditing algorithms, as well as, on a project evaluating the impact of criminal records on employment. Professors Ziewitz and Barocas collaborated on the field-defining ‘Governing Algorithms’ conference at New York University in 2013 and published a widely cited discussion paper on the topic (Barocas et al., 2013). In addition to the core team, we expect that this ISS Collaborative Project will be of interest to a range of other scholars at Cornell, including James Grimmelman (Cornell Tech/Law), Giles Hooker (Statistical Science), Jon Kleinberg (Computer Science), Karen Levy (Information Science/Law), Helen Nissenbaum (Cornell Tech), and Tarleton Gillespie (Communication, Adjunct Associate), as well as to graduate students. Together, we believe that we will make a lasting contribution to the social scientific study of inequality in algorithmic systems at Cornell and beyond.
References


Goodman, J. (2018). *The Legal and Ethical Minefield of AI: Tech has the power to do harm as well as good.* The Guardian.


BIOGRAPHICAL SKETCH

NAME: Wells, Martin T., PhD

eRA COMMONS USER NAME (credential, e.g., agency login): mtwells

POSITION TITLE: Charles A. Alexander Professor in Statistical Sciences

EDUCATION/TRAINING

<table>
<thead>
<tr>
<th>INSTITUTION AND LOCATION</th>
<th>DEGREE</th>
<th>Completion Date</th>
<th>FIELD OF STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>California State University</td>
<td>BS</td>
<td>1983</td>
<td>Mathematics/Psychology</td>
</tr>
<tr>
<td>University of California,</td>
<td>MA</td>
<td>1983</td>
<td>Mathematics</td>
</tr>
<tr>
<td>University of California</td>
<td>PhD</td>
<td>1987</td>
<td>Mathematics/Statistics</td>
</tr>
</tbody>
</table>

A. Personal Statement

Wells is the Charles A. Alexander Professor in the Statistical Sciences, holds a faculty appointment in the Departments of Statistical Science, Social Statistics, Biological Statistics and Computational Biology, Biostatistics and Epidemiology (Weill Medical College), and Cornell’s Law School. Wells has over 200 publications, is a fellow of the major statistics societies, has been editor and associate editor of some of the top tier statistics journals, and has chaired over 30 PhD theses. He has been the primary biostatistician on over $10M NIH grants at Weill Medical College, and been funded by NSF and NIH for methodological research. His legal scholarship has been cited by US Supreme Court numerous times and written about in the popular press. Wells has testified before Congress and has served as an epidemiological/statistical expert and court appointed witness in a number of mass tort cases. Wells has worked in a number of research areas including topics quite relevant to modern event history analysis, demography, and policy analysis, as well as in meta-analysis, spatial statistics, Bayesian analysis, decision theory, multivariate analysis, and high dimensional regression modeling.

B. Positions and Honors

Positions and Employment

2000-          Charles A. Alexander Chair of Statistical Sciences, Cornell University
1993-1999      Associate Professor, Department of Social Statistics, Cornell University
1987-1993      Assistant Professor, Department of Social Statistics, Cornell University

Present Positions

2015-          Faculty Director of the Survey Research Institute, Cornell University
2011-          Professor, Department of Biostatistics and Epidemiology, Weill Cornell Medical College
2000-          Professor, Biological Statistics and Computational Biology, Cornell University
1998-          Professor, Department of Social Statistics, Cornell University
1998-          Professor, Cornell Law School, Cornell University

Other Positions and Honors

Fellow of the American Statistical Association, Institute of Mathematical Statistics and the Royal Statistical Society

Elected Member of the International Statistical Institute


Editor for Journal of Empirical Legal Studies 2003-present

Editor for Statistical Science 2006-2011

Editor for Journal of Multivariate Analysis 2008-present

Member of the National Academy of Science Advisory Panels to evaluate Census

Member of the American Statistical Association’s Advisory Committee for Science Policy
C. Contribution to Science

1. Model Selection: Model selection has had an important place in scientific research because of the necessity to construct parsimonious statistical models. My work, with various colleagues, examines the model selection problem for a number of complex modelling scenarios. Articles (a.-c.) give a unified framework for selecting parsimonious statistical models in particular setting and examines the theoretical properties of a general class of model section rules. Article d. examines various theoretical properties of model selection procedures.

2. Statistical Methods for Genomic Data: Traditional regression problems typically involve a small number of explanatory variables and an analyst can make educated decisions as to which ones should be included in the regression model, and which should not. However, the new age of high speed computing and technological advances in genetics and molecular biology, for example, have dramatically changed the modeling and computation paradigms. It is common practice to use linear regression models to estimate the effects of hundreds or even thousands of predictors on a given response. These modern applications present major challenges. First, there is the so-called 'large p, small n' problem, since the number of predictors, p, e.g. gene expressions from a microarray or RNASeq experiment, often greatly exceeds the sample size, n. Methods controlling the experiment-wise false discovery rate in one predictor at a time analyses often result in few or no discoveries. Second, the model space is huge. For example, for a modest QTL study with 1000 markers, there are $2^{1000}$ possible models. This renders exhaustive search-based algorithms impractical. In the listed articles listed we propose an empirical Bayes, model-based approach to genomic data analysis implement via a fast algorithms.
   d. "A mixture model approach for parallel testing for unequal variances." (with Bar H, Booth J), Statistical Applications in Genetics and Molecular Biology, 11(1) Article 8.

3. Statistical Theory: In general, the theory of statistics is about the mathematical modeling of observable phenomena, using stochastic models, and about analyzing data: estimating parameters of the model and testing hypotheses. In these contributions, we study various estimation and testing procedures, considering their theoretical properties and investigate various notions of optimality for complex statistical models that have practical applications.
4. **Statistical Methods for Complex Biological Modeling:** Biological systems study interactions between the biology and behavior at all levels of biological organization from the small scales up to the large scales of populations and communities. Mathematical modeling of complex systems is wide and varied. Each of the articles develop inferential procedures for particular complex biological models.

5. **Collaborations in randomized trials:** I have extensive experience in collaborations with investigators involving randomized trials.

D. **Research Support**

**Ongoing Research Support**

<table>
<thead>
<tr>
<th>Fund ID</th>
<th>Principal Investigator</th>
<th>Proposal Title</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>U19AI111143-02</td>
<td>Nathan</td>
<td>NIH NIAID Center for Drug-Resistant Tuberculosis: Hits, Leads and Targets</td>
<td>7/1/2014</td>
<td>6/30/2021</td>
</tr>
<tr>
<td>DMS-1611893</td>
<td>Wells</td>
<td>Directorate for Mathematical &amp; Physical Sciences NSF</td>
<td>8/15/2016</td>
<td>7/31/2019</td>
</tr>
</tbody>
</table>

Center for Drug-Resistant Tuberculosis: Hits, Leads and Targets
The goal of this study is to address significant gaps in understanding how TB bacteria establish and maintain infection in the human body as well as how agents might attack vulnerabilities in the bacteria.
Role: Statistician

Variable Selection When P>> N: Beyond the Linear Regression and Normal Errors Model
The goal of this project is to develop new statistical models and computational algorithms for high dimensional, low sample size, high-throughput biological data, including new methods for the analysis of microarrays, the identification of quantitative trait loci, association mapping, label-free shotgun proteomics and metabolomics. The proposed methods involve innovative extensions of modern statistical building blocks, including the use of random effects for regularization, shrinkage estimation, Bayesian statistics, and mixtures for posterior classification and prediction.
Role: PI
IFEOMA AJUNWA, J.D., PH.D.

ORGANIZATIONAL BEHAVIOR DEPARTMENT
CORNELL UNIVERSITY ILR SCHOOL
396 IVES HALL, ITHACA, NY 14853

IAJUNWA@CORNELL.EDU
(917) 291-1422
WWW.IFEOMAAJUNWA.COM

EDUCATION
2017 Columbia University, Ph.D., Sociology
2012 M.Phil., Sociology (earned en route to Ph.D.)
2007 University of San Francisco, J.D., Law
2003 University of California, Davis, B.A., Political Science and Sociology

CURRENT ACADEMIC APPOINTMENTS
2017- Cornell University ILR School, Assistant Professor
2017- Cornell Law School, Faculty Associate Member
2017- Center for the Study of Inequality, Faculty Affiliate
2017- Faculty Associate, Berkman Klein Center for the Study of Internet and Society

PUBLICATIONS
2018 The Quantified Worker (under contract with Cambridge University Press, forthcoming book)
2018 The Formerly Incarcerated’s Quest For Equal Protection in the Labor Market, (with Angela Onwuachi-Willig), 116 Nw. L. Rv. 6.
2017 Limitless Worker Surveillance, 105 Cal. L.Rev.736 (with Kate Crawford and Jason Schultz)
2017 “Big Data Risks of Workplace Wellness Programs and How to Mitigate Them,” Harvard Business Review
2016 “Health and Big Data: An Ethical Framework for Health Information Collection By Corporate Wellness Programs,” The Journal of Law, Medicine & Ethics, 44 (with Kate Crawford and Joel Ford)
2016 Genetic Data and Civil Rights, 51 Harv. C.R.-C.L. L. Rev. 75 (2016)


**Publications in Progress**

*Hiring By Algorithm* (law review article, in progress)

*Automated Hiring Platforms as Technological Intermediaries and Digital Brokers* (peer-review article in progress)

*The Unbounded Workplace: The Ethics of Wearables and Productivity Apps In the Workplace* (peer review, invited to submit, Journal of Business Ethics).

**Other Select Publications**


2017  *Corporate Surveillance is Turning Human Workers Into Fungible Cogs*, The Atlantic, May 19th

2015  *There is No Guarantee of Anonymity for Online Genetic Databases*, New York Times Room for Debate, March 4th

2014  *Do You Know Where Your Health Data Is?*  Huffington Post Impact

**SELECT INVITED PRESENTATIONS AND CONFERENCES**

2017  Tech Progress and Algorithmic Age Discrimination, UC Berkeley Law, November

2017  Mind to Mind, *Data Discrimination and Algorithms*, Stanford University, October

2017  Plenary Speaker, *Automated Hiring Platforms*, WORK2017, Turku, Finland, August

2017  *Technology and the Self*, Workshop at Academy of Arts and Sciences, Cambridge, Massachusetts, May

2017  *The Quantified Worker*, Harvard Law School, Cambridge, Massachusetts, May


2017  *Wellness or Else: Workplace Wellness Programs*, SXSW, Austin, Texas, March

2017  *Algorithmic Bias: Where It Comes From And What to Do About It*, LibrePlanet, Cambridge, Massachusetts, March

2017  *Health Information Outside of a Clinical Setting: PRIM&R Workshop on Human Subject Research*, Boston, Massachusetts, March
2016  
*Big Data and Employment*, Equal Employment Opportunity Commission (EEOC), Washington, D.C., October

2016  
*A Contractual Escape Clause For the Third Party Doctrine*, New Scholars Workshop; Panel on New Forms of Scholarship; Discussion Group, Wellness Programs, Southeastern Association of Law Schools (SEALS) conference, Amelia Island, FL, August

2016  
*Hiring by Algorithm*, Lutie Lytle Writing Workshop at University of Iowa Law School, Iowa City, IA, July

2016  
Closing Keynote Lecture, Health Privacy Rights Summit, Georgetown University School of Law, Washington, D.C., June

2016  
Discussant, “Risk, Liability, & Norms in a Digital Age,” Law and Society Annual Meeting, New Orleans, LA, June

2016  
*Limitless Worker Surveillance*, at the Privacy Law Scholars Conference, George Washington University Law School, D.C., June

2016  
*Biased Data, Biased Decisions*, Consumer Financial Protection Bureau (CFPB), Washington, D.C., May

2016  
*Hiring by Algorithm*, Google Re:Work Conference, Mountain View, CA, April

2016  
*Hiring By Algorithm*, Unlocking the Black Box, Yale ISP Conference, New Haven, CT, April

2015  
*Limitless Worker Surveillance* – Microsoft Research, New York, NY, July

2015  
*Genetic Data and Civil Rights* – This Week In Health Law Podcast with Professors Frank Pasquale and Nicholas Terry, July

2015  
*The Quantified Self* – The Berkman Center for Internet and Society at Harvard University, Cambridge, MA, June

2015  
*Genetic Coercion* – Data & Society Research Institute, New York, NY, June

2015  
*Workplace Wellness Programs In an Era of Quantified Work*, Health Privacy Rights Summit, Georgetown University School of Law, Washington, D.C., June

2014  

**Fellowships, Grants, and Scholarships**

2017  
ILR School Theme Project Grant – Auditing Algorithms ($20,000)

2017  
ILR School Theme Project Grant – Gender and the Tech Industry ($20,000)

2015  
National Science Foundation (NSF) Doctoral Research Improvement Grant

2013  
Spivack Travel Grant to present research at American Sociological Association Conference

2012  
Ford Foundation Predoctoral Fellowship (Honorable Mention)

2011  
Cornerhouse Travel Grant to present research at the American Sociological Association

2010-2015  
Paul F. Lazarsfeld Fellowship, Columbia University (full tuition and stipend for 5 years)

2006-2007  
American Association of University Women (AAUW) Selected Professions Fellowship for last year of law school study

**Relevant Teaching:** Technology, Management, and the Law, ILROB 5230, Cornell ILR School
Solon Barocas  
Department of Information Science  
Cornell University  

Phone: 646-456-9663  
sbarocas@cornell.edu  

a. Research Program  

Dr. Barocas’ research places concerns about the growing role that novel computational techniques play in decision-making on firm empirical ground by interrogating how, at a technical level, these practices implicate core values like privacy, fairness, and accountability. To date, he has explored issues of fairness in machine learning, methods for bringing accountability to automated decisions, the privacy implications of inference, and the relationship between privacy and economic inequality. His work has appeared in *University of Pennsylvania Law Review* and *California Law Review*, among other places, and has informed policy documents on these topics. His research ultimately aims to make computational techniques less inscrutable, more accountable, and fairer through a mix of technical, institutional, and policy mechanisms.

b. Professional Preparation  

<table>
<thead>
<tr>
<th>Institution</th>
<th>Degree</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown University</td>
<td>International Relations</td>
<td>B.A.</td>
</tr>
<tr>
<td>London School of Economics</td>
<td>Modern Culture and Media</td>
<td>B.A.</td>
</tr>
<tr>
<td>New York University</td>
<td>International Relations</td>
<td>M.Sc.</td>
</tr>
<tr>
<td></td>
<td>Media, Culture, and Communication</td>
<td>Ph.D.</td>
</tr>
</tbody>
</table>

b. Appointments  

- **2017-**  
  **Assistant Professor**, Department of Information Science, Cornell University  
- **2016-2017**  
  **Visiting Assistant Professor**, Department of Information Science, Cornell University  
- **2016-2017**  
  **Postdoctoral Researcher**, Microsoft Research  
- **2014-2016**  
  **Postdoctoral Research Associate**, Center for Information Technology Policy, Princeton University  

c. Products  

**PRODUCTS MOST CLOSELY RELATED**  


OTHER PRODUCTS


d. Synergistic Activities

Brooke Erin Duffy

CONTACT INFORMATION
478 Mann Library Building
Cornell University
Ithaca, NY 14853
Office: 607.255.8403
Email: bduffy@cornell.edu
Website: www.brookeduffy.com

EDUCATION
Ph. D., Annenberg School for Communication, University of Pennsylvania, 2011
B. A. (honors & highest distinction), College of Communications, The Pennsylvania State University, 2002

CURRENT APPOINTMENT
Cornell University
Assistant Professor, Department of Communication, 2016-
Faculty Affiliate, Center for the Study of Inequality; Member, Media Studies Initiative

PRIMARY RESEARCH AREAS
Media, technology, and society; gender, feminism, and inequality; work and labor in the digital age; cultural production and online self-presentation

PUBLICATIONS: BOOKS


PUBLICATIONS: JOURNAL ARTICLES


UNDER REVIEW


INVITED TALKS
University of Southampton, UK, April 2018
Critical Cultural Seminar, University of Minnesota-Twin Cities, March 2018
McLuhan Center for Culture and Technology, University of Toronto, January 2018
Annenberg Research Seminar, University of Southern California, November 2016
University of North Carolina-Chapel Hill, March 2016
Columbia University Graduate Communications Colloquium, March 2015
Great Works Symposium on Media, Drexel University, June 2015
University of Georgia, March 2014.
Social Thought Program Lecture in Communications, November 2013.

CONFERENCE PRESENTATIONS


HONORS & AWARDS
Outstanding educator for having influenced a Merrill Presidential Scholar, 2017
Visiting Honors Scholar, Media & Communication, Muhlenberg College, Fall 2015
Emerging Scholar Award, Critical/Cultural Studies, National Communication Association, 2014
Lillian Lodge Kopenhaver Outstanding Woman Junior Scholar Award, 2014
Temple University Merit Award for Research: 2012, 2013, 2014
Annenberg Dissertation Research Fellowship, 2010
Most Innovative Poster Award, National Communication Association, 2010
Full tuition and research stipend, Annenberg School for Communication (2005-2010)

GRANTS
Institute for the Social Sciences Grant ($10,967), Cornell University, 2017-2018
The Waterhouse Family Institute for the Study of Comm. and Society Grant ($6,000), 2014-2015
Organization for Research on Women in Communication Grant ($1,000), 2014
Summer Faculty Research Award, Temple University ($7,000), 2014
Summer Faculty Research Award, Temple University ($7,000), 2013
Diamond Research Scholar Faculty Mentor Stipend, Temple University ($500), 2012
Annenberg Oxford Institute on Global Media Policy (fully funded), 2007

RELEVANT TEACHING
Media, Technology, & Society (Doctoral seminar, Spring 2018)
Gender and Media (Undergraduate seminar, Fall 2017)
New Media & Society (Undergraduate lecture, Spring 2017)
Media and Cultural Production in the Digital Age, (Doctoral seminar, Temple University, Spring 2016)
Qualitative Research Methods (Doctoral seminar, Temple University, Spring 2015)
MALTE ZIEWITZ

Department of Science & Technology Studies                        mcz35@cornell.edu
Cornell University                                            +1 (917) 214-9732
313 Morrill Hall, Ithaca, NY 14853                               http://zwztz.org

ACADEMIC POSITIONS

2014- Assistant Professor, Mills Family Faculty Fellow
Department of Science & Technology Studies, Cornell University
Graduate field membership: Science & Technology Studies, Information Science
    Faculty Fellow, David R. Atkinson Center for a Sustainable Future (2017-)
    Faculty Fellow, Cornell Institute for Healthy Futures (CIHF) (2016-)
    Keeton House Fellow, Cornell University (2015-)
    Associate Member, Centre for Science, Knowledge, and Policy (SKAPE), University of Edinburgh (2014-)

2012-2014 Postdoctoral Research Fellow
Department of Media, Culture, and Communication, New York University
Associated Fellow, Institute for Information Law, New York University

EDUCATION

2012 D.Phil., Management Studies (S&TS Group), University of Oxford
    Supervisor: Steve Woolgar, Examiners: Javier Lezaun, Annemarie Mol
2006 M.P.A., Public Administration, Kennedy School of Government, Harvard University
2003 First State Exam, Law (J.D. equivalent, GPA: 14.75, ‘sehr gut’), University of Hamburg
2001 Erasmus Exchange Fellow, Sapienza Università di Roma, Italy

SELECTED GRANTS AND FELLOWSHIPS

Under review CAREER Grant, National Science Foundation, $429,785, PI
“Understanding and Advancing User Participation in Algorithmic Systems.”

2017 Small Grant, Institute for the Social Sciences, Cornell, $9,000, PI
“Restoring Credit: How Users Understand and Interact with Credit Scoring Systems.”
    Co-PI: Ranjit Singh

2016 Workshop Grant (#1656276), National Science Foundation, $49,999, Co-PI
“Technoscientific Constitutionalism: Exploring New Horizons in STS Research at the Intersection of Science, Technology, Law, and Governance.” PI: Ben Hurlbut,
    Co-PIs: Christopher Kelty, Shobita Parthasarathy

2014 Symposium Grant, Foundation for the Sociology of Health & Illness, £3,000, Co-PI

2011 Research Grant, Saïd Business School Student Support Fund, £1,000

2009  Workshop Support, Networks for Web Science, Engineering and Physical Sciences Research Council (EPSRC) (with Christian Pentzold), £3,000

2008  PGP Corporation Scholarship, University of Oxford, £11,000

2007  Doctoral Scholarship, German Academic Exchange Service (DAAD), €21,600

2004  McCloy Fellowship, Harvard University/German Academic Merit Foundation, $105,650

ERP Fellowship, German Academic Merit Foundation (declined)

SELECTED PUBLICATIONS

— Edited Collections


— Journal Articles


— Book Chapters


— Invited Essays, Discussion Papers, Research Reports


**SELECTED SCHOLARLY PRESENTATIONS**


“Algorithmic Power,” Power Switch Symposium, CRASSH, University of Cambridge, Cambridge, United Kingdom, March 31, 2017


“Enginology,” Workshop on Revelation, Linköping University, Vadstena, Sweden, April 8-9, 2016

“Spectacular Algorithms,” CSTMS Colloquium Series, Center for Science, Technology, Medicine, and Society, UC Berkeley, November 5, 2015


“Accounting for Rankings: Notes from an Ethnography with SEO Consultants,” Intel Science & Technology Center (ISTC) for Social Computing, Georgia Tech, Atlanta, GA, May 21, 2014

“Unscaling Ethnography,” Big: Culture and Data in the Digital Field, UC Irvine, CA, April 11, 2014


“Enacting Governance,” Social Media Governance workshop, Faculty of Law, University of Haifa, Haifa, Israel, July 24, 2013

