# Establishing Sex
# The Scientific Quest to Support a Controversial Binary

by

Jeffrey W. Lockhart

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Sociology)
in the University of Michigan
2022

Doctoral Committee:

Associate Professor Elizabeth E. Bruch, Co-Chair
Associate Professor Erin A. Cech, Co-Chair
Professor Jason D. Owen-Smith
Associate Professor Joy Rohde

Jeffrey W. Lockhart

jwlock@umich.edu

ORCID iD: 0000-0003-1519-9588

## Acknowledgments

This dissertation would not have been possible without myriad sources of funding and resources, small and large, that I have had access to over the last seven years. These include an NICHD training grant to the Population Studies Center at the University of Michigan (T32HD007339), the University Library's all-access subscriptions to databases such as the Web of Science, the University's Advanced Research Computing infrastructure, Elizabeth Bruch's various budgets, and too many smaller sources to name, each of which was nevertheless important.

Likewise, the intellectual debts I have accrued are too many to enumerate, but they surely include Elizabeth Armstrong, Will Beischel, Chalem Bolton, Jamie Budnik, Dom Canning, Megan Carroll, Anne Clarke, Sarah Fenstermaker, Elly Field, Connor Gilroy, Judith Green, Jack Halberstam, Dan Hirschman, Abigail Jacobs, Nora Johnson, Katrina Karkazis, Greta Krippner, Margo Mahan, Christin Munsch, Os Keyes, Madi Pape, Sarah Richardson, Oliver Rollins, Gayle Rubin, Kouji Takahashi, Mike Thompson-Brusstar, Louie Dean Valencia, Sari van Anders, and Jesse Yeh. A number of people deserve special thanks. Gary Weiss introduced me to the world of research and gave me both the freedom and support to grow into it. Orit Avishai introduced me to sociology not simply as an academic field, but as a praxis that could shape the world for good, a lesson I must return to more often. Lauren Wilcox introduced me to the issues at the heart of this dissertation, a seed that got under my skin for years before I finally let it blossom. Elizabeth Bruch supported and believed in me as a scholar even before I had proven myself and even when my work seemed like nonsense. Erin Cech never failed to hold my work to the same high standards she does her own, resulting in the most constructive critiques I ever received. Jason Owen-Smith shared his encyclopedic knowledge of STS debates and always encouraged me to follow my own path through them. Joy Rohde taught me more, minute-for-minute, about good historical claims-making than anyone else could.

I would not have been able to complete this dissertation and degree without personal and professional support of many other kinds, within and without the discipline. I found much of this in feminist, queer, and labor communities like Aút, Caucus, the GenderSci Lab, GEO, LIVE, NeuroGenderings, SWS, and the Sexualities Section. Some of the individuals who guided me, sustained me, and helped me sustain myself most are Abby Azari, Chalem Bolton, Tristan Bridges, Cati Connell, Shauna Dyer, Liz Ela, Tay Field, Connor Gilroy, Jax Gonzalez, Theo Greene, Alex Ingraham, Abbie Jacobs, Chelle Jones, Yael Kenan, Parker Koch, Greggor Mattson, Jay Orne, Lynette Shaw, Mike Thompson-Brusstar, Alex Thum, Alejandro Vigo Carmago, Dan Walden, and Pat Zabawa. And, of course, Donna.

## Table of Contents

**List of Figures**

## Abstract

Sex and gender lie at the heart of a vast array of social, cultural, and political institutions and controversies. From highly politicized issues like transgender people's access to bathrooms and women's under representation in science, to more mundane ones like the color of children's toys, sociologists have long recognized sex as a "master category" that structures every other dimension of life. Yet social change and human diversity challenge the essential, categorical understanding of sex that underpins the use of sex in this way. Biology has been a key reference point and battleground for contesting the sex binary. This dissertation builds on the sociology of science and feminist science studies using a mix of computational and qualitative methods to critically examine how scientists work to establish the sex binary as the authoritative ontology of sex.

I examine three key sites in the process of the scientific production of sex categories. These sites—production of science in labs, presentation of science in writing, and reception of papers in scientific fields—are each important contexts where scientists do the work of establishing sex, its boundaries, contents, and meanings as a category system.

Chapter Two is built on technographic analysis of all 53 research papers using machine learning to predict sex from brain scans, along with their supplemental materials, code, comments, replies, and retraction notices. Using this genre of sex research as a case study, I demonstrate how machine learning methods like support vector machines have unique affordances that can both help and hinder efforts to establish the binary nature of sex. Whether they ultimately do depends on researcher choices and intent. This poses a key intervention in the sociology of algorithms by theorizing how they can be deployed to resist (rather than enact) social change.

Chapter Three draws on the literature reviews of 387 books and articles about sex differences to argue that the scientific authority of binary sex is often rhetorically established through revisionist histories of the field. I outline three commonly deployed revisionist frames: that binary sex is uncontested, that its proponents are Galilean martyrs fighting dogma, and that its opponents are child abusers.

In Chapter Four I merge newly available NSF data on the demographics of all PhDs in the US since 1970 with Web of Science data on 69 million academic publications. Drawing on my qualitative work, I develop and validate an automated system for categorizing how life science research discusses sex. In keeping with theoretical and small-scale empirical work, I demonstrate that the more sex essentialism biology and health subfields publish, the fewer women they will grant PhDs to in future years. The inverse also holds: more anti-essentialist, feminist biology corresponds to more women PhDs in future years. This demonstrates that theories of gendered occupational segregation, which typically focus on variation in occupational requirements and norms, are incomplete without attention to variation in beliefs and scientific claims about gender.

Taken together, this dissertation argues that scientists' quest to establish (or undermine) the sex binary influence every stage of scientific production, including design, interpretation, presentation, and even the composition of the future scientific workforce. Sex is a powerful factor that, beyond being shaped by science, actively shapes science.

# Chapter 1 Introduction

At the core of this dissertation is a simple question: How does biological sex essentialism persist in science? Sex is one of the most fundamental social categories and also one of the most heavily biologized today (Epstein 2007; West and Zimmerman 1987). This differentiates sex from most contested scientific concepts, which rarely enter the public consciousness. It is thus imperative to understand how sex, a social category system[1], is biologized and made scientific, both because of its social consequences and because it operates differently than other scientific controversies. Science and technology studies has a venerable tradition of exploring the social production of scientific facts in laboratories and field sites. Such work demonstrates how data and observations come through scientific cultures and practices to become evidence and facts (e.g. Latour 1983). Simultaneously, much work in the sociology of science has explored how the social conditions of science—the politics, networks, funding, incentives, and so on—shape what what questions are asked and which answers are accepted (e.g. Epstein 1996; Panofsky 2014; Waidzunas 2012).

In this dissertation I embark on a third kind of analysis, drawing on the concerns of both. While I take the fundamentally cultural and political object of sex categorization as my object of study, I bring to it an analysis of laboratory production. I aim to show *how* (essentialist, binary) sex is established in the content of scientific work. Apart from the obvious social incentives for scientists to produce normative understandings of sex, what do scientists actually *do* in their design, analysis, and writing to reproduce this highly consequential social category system and its concomitant inequalities? This approach places my dissertation squarely in the tradition of feminist science studies, which is traceable as a collective movement back to the mid-1970s. I examine three key sites in the process of the scientific production of sex categories. These sites —production of science in labs, presentation of science in writing, and reception of papers in scientific fields—are each important contexts where scientists do the work of establishing sex, its

[1] Here I do not mean the union of gametes that produces offspring in sexually reproducing species including people, peonies, and piranha. Instead, I mean the broader notions of 'maleness' and 'femaleness', of men and women's 'natures', and the myriad physical, psychological, and social characteristics that are said to neatly align with them. As a category system for living organisms, especially people, sex is very much socially constructed, with origins beyond science (Laqueur 2001).

boundaries, contents, and meanings as a category system. There is no single place or process by which sex seeps into science. Instead, as a pervasive cultural idea, sex manifests in myriad quotidian moments and processes throughout science. Understanding these processes teaches us both about sex categorization specifically, and about the production of scientific knowledge and social categories generally.

Sex is not the only fundamental social category system—what some theorize as a "master category" system, influencing all other aspects of identity, life, and politics (West and Zimmerman 1987)—nor is it the only one to be biologized. Race, class, sexuality, disability, even political party are at different times and places the most salient, intersecting category systems governing social life (Collins [1990] 2000; Weeks 1990). All have been biologized and subjected to scientific study (which is to say, scientific [re]production), but none today involves such a widespread, essentialist scientific paradigm as sex does (Epstein 2007). This makes it a key case for studying the scientific production of human social categories, which has for centuries been a primary means of "making up people" and exerting social control over them because of science's epistemic authority (Daston and Galison 2007; Foucault 1990; Hacking 1986; Igo 2007). Categories, in turn, are the fundamental bases of and objects of all social struggle (Becker [1963] 1973; Bourdieu 1984; Geoffrey C. Bowker and Star 1999; Durkheim and Mauss [1903] 1967; Goffman [1963] 1986). To understand any sociological object, then, we must understand how categories are made and contested.

The essentialist view of sex is dominant in science, but not because it is the best representation of available evidence. For over half a century feminist scientists and epistemologists have demonstrated the inadequacy of scientific explanations based on a strict binary categorical understanding of sex (Fausto-Sterling 2000; Hubbard 1990; Richardson 2013). Nevertheless, this paradigm of sex manifests in over a thousand papers per year. There multiple journals, professional societies, and even a US federal mandate affecting $41.7 billion in annual research funding, dedicated to advancing an essentialist, binary understanding of sex.

The essentialist understanding of sex in science has far reaching effects. It led the Food and Drug Administration to wrongly prescribe different medication dosages for men and women (Richardson et al. 2015). Claims about the biological essence of sex have led politicians both abroad and in the US to try and shutter gender studies programs in universities, to sponsor (and sometimes pass) hundreds of pieces of legislation curtailing the rights and freedoms of LGB and especially trans and intersex people, and more. Indeed, the 'biological reality' of sex is at the center of the modern culture war in numerous countries around the world (Ferber and Butler 2020). As sociologists have long demonstrated, sex categorization is fundamental and routine social process pervading every aspect of social life, constantly negotiated and established among people, whether they are lay, doctors, or scientists (Almeling 2020; Davis 2015; Garfinkel [1967] 2006; Martin 1991; Pape 2021; West and Zimmerman 1987).

❦❦❦

Methodologically, my scholarship in this dissertation and beyond brings together both computational, "big data," approaches and qualitative, archival ones. Traditionally these styles of analysis are conducted by separate groups of scholars, often in separate venues and even separate citation circles. This makes it challenging to bring them together. Typical approaches use one as a second-fiddle supplement to the other. Either a quantitative analysis is leverage to show that some qualitative conclusion is widespread, or a qualitative analysis is used to add narrative detail and thus believably to a primarily quantitative one. Each of these styles of analysis, on their own and in their two primary combinations, has yielded valuable sociological insight.

There is, however, an alternative way to bring them together that opens up the possibility of asking a new set of questions. For example, the deep knowledge of the case in my dissertation —sex differences research—that I gained from sustained qualitative analysis of its rhetoric enabled me to design and validate quantitative measures of specific scientific paradigms in Chapter Four. Paradigms have not traditionally been objects of quantitative study because they do not exist as variables in any of the data sets used by science of science researchers. Even the highly detailed JEL code system in economics focuses on objects of study and methods, rather than sensibilities or arguments. On the other extreme, the clustering, network, and embedding approaches increasingly popular in large scale science of science are content-agnostic and therefore unable to trace the contours of paradigmatic conflicts that we know are of substantive importance, unless those conflicts manifest as fairly radical schisms in writing style or citation networks. When it comes to scientific struggles over sex, the conflict is not structured as such a schism, and purely quantitative approaches render it invisible, limiting not just the scope of cases that can be studied, but the scope of ways we can imagine and theorize scientific conflict.

Chapter Two offers another example of how a union of big data and qualitative approaches can open up new questions. Quantitatively-minded approaches to the question of group categorization in machine learning overwhelmingly focus on the algorithms' performance: here we have the numerous studies examining the distributions of errors to make arguments about how this or that group is disadvantaged by an algorithm. Meanwhile, numerous qualitative approaches focus on the interpretations and social effects of algorithms' use, or the social meanings encoded in the data the algorithms draw on. By investing in both simultaneously, I am able to explore how the inner workings of algorithms interplay with the meanings and social effects in and around them. More directly, many important questions about technical systems do not require understanding of how they actually operate, and many others do not require understanding the social milieu in which they are used, but some questions cannot be understood without both. In a way, this is an unusually realist position for me: the technical/quantitative substance of things places important constraints and opportunities on interpretations and effects that we need to attend to. Yet as I show in Chapter Two, scientists' attempts to grapple with the limits and capitalize on the opportunities presented by algorithms are anything but determined by them. Materiality affects routes more than ends.

❦❦❦

The rest of the dissertation proceeds in three parts: Chapter Two, Doing the Work; Chapter Three, Presenting the Work; and Chapter Four, Experiencing the Work. In Chapter Two, I take a close look at exactly how scientists manufacture a sense of binary sex out of the much more complicated reality. In order to do so, I conduct technographic (Bucher 2016) analysis of all 53 research papers using machine learning to predict sex from brain scans, along with their supplemental materials, code, comments, replies, and retraction notices. Using this genre of sex research as a case study, I demonstrate how machine learning methods like support vector machines have unique affordances that can both help and hinder efforts to establish the binary nature of sex. Whether they ultimately do depends on researcher choices and intent. This poses a key intervention in the sociology of algorithms by theorizing how they can be deployed to resist (rather than enact) social change.

In Chapter Three I turn from the production of binary sex in research to its presentation. Scientific writing is not only about the presentation of specific facts from a single study. The narrative of how those results fit into the broader set of knowledge and literature on sex is key for credibility and interpretation. Drawing on the literature reviews of 387 books and articles about sex differences, I argue that the scientific authority of binary sex is often rhetorically established through revisionist histories of the field. That is, scientists misrepresent past work in ways that make their present work seem more legitimate. I outline three commonly deployed revisionist frames: that binary sex is uncontested, that its proponents are Galilean martyrs fighting dogma, and that its opponents are child abusers.

In Chapter Four I turn from the presentation of sex research to its reception in the scientific community. Once scientific research is conducted and written up, it is sent to editors and reviewers who either reject it or shape it until they believe it is acceptable, thereby reflecting, shaping, and enforcing the norms of their discipline, and granting published work their imprimatur. Thus published work represents not only the authors' beliefs about what kinds of claims about sex are legitimate, but also those of their fields more broadly. And by virtue of being published, some claims become more legitimate. Since biological sex research is often about the relative abilities and preferences of men and women, and scientists are constantly evaluating one another based on abilities and preferences, we may expect that scientists' beliefs about sex influence the gender ratio of their fields. That is exactly what I find. I merge newly available NSF data on the demographics of all PhDs in the US since 1970 with Web of Science data on 69 million academic publications. In keeping with theoretical and small-scale empirical work, I demonstrate that the more sex essentialism biology and health subfields publish, the fewer women they will grant PhDs to in future years. The inverse also holds: more anti-essentialist, feminist biology corresponds to more women PhDs in future years. This demonstrates that theories of gendered occupational segregation, which typically focus on variation in occupational requirements and norms, are incomplete without attention to variation in beliefs about gender.

Taken together, this dissertation argues that scientists' quest to establish (or undermine) the sex binary influence every stage of scientific production, including design, interpretation, presentation, and even the composition of the future scientific workforce. Sex is a powerful social category system that, beyond being shaped by science, actively shapes science.

## Chapter 2 Doing the Work: Machine Learning, Neuroimaging, and Epistemologies of Sex

Algorithms have "made inroads in almost every major social institution," including media consumption, teacher evaluations, insurance pricing, and policing (Burrell and Fourcade 2021:1). To date, the literature on algorithms and society has largely organized into two positions. One strand theorizes how algorithms are used to consciously and fundamentally transform society (Burrell and Fourcade 2021; Cheney-Lippold 2017; Kiviat 2019). The other strand theorizes how algorithms tend to unreflexively reproduce longstanding social patterns (Benjamin 2019; Eubanks 2017; Noble 2018). These positions are not necessarily contradictory. An algorithm may fundamentally transform, say, Medicaid delivery by removing caseworkers and appeals—the conscious intent—while simultaneously reproducing centuries old conceptions and deprivations of the "undeserving poor"—an unexamined theory[2] of wealth inequality (Eubanks 2017). Indeed, replacing theoretical reflection or justification with 'data' is often presented as a virtue of algorithms (Bowker 2014). But theory is omnipresent in the measuring and modeling that algorithms operate on, whether practitioners are aware of it or not. Algorithmic fairness scholars and classic sociology alike have identified unexamined and mismatched theory in algorithmic and quantitative systems as a primary mechanism by which methods reproduce social structures (Cicourel 1964; Jacobs 2021; Jacobs and Wallach 2021; Westbrook and Saperstein 2015).

Both strands of theorizing about algorithms have been limited by their focus on *algorithms for production*: algorithms that either act in the world (e.g., by choosing who gets an interview or sees a housing advertisement), or directly guide action (e.g., telling a judge, doctor, case worker, or drone pilot whether a specific person is 'high risk'). Beyond this, algorithms are increasingly used *for discovery.* That is, researchers across a wide array of academic fields use algorithms to better understand the world, as in computational social science (Azari et al. 2020; King et al. 2009; Nelson 2021). Algorithms for discovery may use the same code and data as

[2] Throughout, I use "theory" in the sense of "an understanding of the world." Sometimes this is explicit, as in a scientist's theory of how something works or an activist's theory of change. Other times it is implicit. These are not fundamentally different kinds of theory—the theory of undeserving poor has been both to different people at different times. The literatures I draw on are concerned with transitioning theories from implicit to explicit.

their production counterparts, but their output is not directly used to guide action. Instead researchers interpret their output and use it to make knowledge claims (e.g. writing a paper about the factors that predict recidivism). In discovery contexts, unlike production contexts, measurement and theory evaluation are the point of using algorithms. Such algorithms are a test of whether thoughtful, explicit attention to theory and measurement can prevent algorithms from reproducing harmful social patterns.

Algorithms for discovery are particularly influential in modern neuroscience, where neuroimaging data (brain scans) are increasingly 'big data' processed with machine learning algorithms. Among myriad traits studied, sex has emerged as the 'gold standard' for machine learning with neuroimaging data. Sex is used both as an object of study (scientists try to understand the nature of sex using machine learning) and as a reliable benchmark for new algorithms (scientists try to show their new algorithmic system works by showing it predicts sex from brains). But the nature and meaning of sex is heavily contested both in society generally and among scientists studying sex (Bluhm 2021; Eliot et al. 2021; Lockhart 2020, 2021; Pape 2021). Algorithms intervening in these debates might either revolutionize our understandings or retrench traditional conceptions. Thus the neuroscience of sex makes an ideal case of algorithms for discovery to test theories of algorithmic systems' effects.

I argue that algorithms can reproduce social structures through conscious, explicit theorizing. I extend the theory of "measurement as governance" in algorithms to argue that while unconsidered mismatches between theory and methodology often account for the way production algorithms reproduce the status quo (Jacobs 2021), algorithmic systems may also reproduce the status quo through highly-considered, highly-aligned theory and methodology, as they do in brain sex research. More directly, discriminating between social groups can be intentional rather than incidental in algorithmic systems. To make this argument, I employ technographic analysis of all 53 academic papers that use machine learning[3] to predict sex from scans of human brains, along with their supplemental materials, code, commentaries, and retraction notices. I peer into machine learning algorithms that have been black boxed and demonstrate how design and interpretation choices can either reinforce or undermine the sex binary—a contested scientific

[3] Most work on algorithms and society examines cases of machine learning, typically supervised classification and regression. I use "algorithms" when referring to the general literature and more specific terms like "machine learning" and "supervised classification" when referring to particular systems.

paradigm holding that sex and gender are best understood as categorical (male/female), with differences driven by innate biology (Keyes, Hitzig, and Blell 2021).

## 2.1 Background

My argument draws on several intellectual traditions. First, I summarize the two major threads of algorithms and society research that I extend. Second, I review the smaller literature on algorithms for discovery and provide background on algorithms in neuroscience. Third, I discuss the sociology of classification and classification of sex specifically. Finally, I give context for my case, brain sex research.

### *2.1.1 Production Algorithms and Society*

Research on algorithms and society has focused on production algorithms: those algorithms used to make or guide individual decisions in the world. Social research on production algorithms is organized in two main threads. In one thread, there is a large literature exploring how algorithms produce social change. Social scientists have theorized the spread of algorithms and their transformational effects on society as fundamentally about changing knowledge and classification. The ubiquitous classification systems in markets produce a "new economy of moral judgement" (Fourcade and Healy 2016:9). Others argue that the adoption of algorithms "produces new versions of the world" (Cheney-Lippold 2017) or, most dramatically, produces "the next human nature" (Zuboff 2018:461).

The other thread emphasizes how algorithms maintain, rather than revolutionize, the status quo. Summing up the resistance to the change narrative, Eubanks pointedly writes, "It is mere fantasy to think that a statistical model or a ranking algorithm will magically upend culture, policies, and institutions built over centuries" (2017:178). This is especially true for machine learning systems, which are designed to find and reproduce patterns in data from the already-existing world. ML is fundamentally conservative, because it cannot anticipate the change and emergent behavior inherent to complex systems like social life (Birhane 2021). To get racist technology, there is "no malice needed, no N-word required, just a lack of concern for how the past shapes the present" (Benjamin 2019:60). Likewise, production algorithms that harm trans people often result not from overt malice but rather from neglecting their existence, or from pro-social attempts to include trans people without rethinking an underlying exclusionary concept of gender (Keyes 2018). Production systems might use "automated technologies to re-inscribe essential notions of difference" between groups like men and women, as an app that screens pictures of faces to decide who is let into its "female only" social network does (Scheuerman, Pape, and Hanna 2021:2). But convincing people there is an essential difference between men and women is not the point of this technology, it is an implicit theory behind the explicit goal of gatekeeping 'female space.'

According to this literature, the beliefs and assumptions of algorithm engineers matter (Seaver 2019), but primarily through their inattention and ignorance. Algorithmic fairness scholars call for conscious attention to measurement, assumptions, data sources, and mismatches between them as solutions for algorithmic injustice (Crawford and Paglen 2019; Jacobs 2021; Jacobs and Wallach 2021). This emphasis on inattention mirrors sociological descriptions of how culture seeps into social scientific methods (Cicourel 1964; Westbrook and Saperstein 2015) as well as Kiviat's (2019) theory of when people will or will not contest algorithmic systems. If the assumptions are not a problem personally for the scientist, engineer, or policy maker, the system passes unquestioned.

### *2.1.2 Algorithms for Discovery*

Much less critical attention has been paid to algorithms for discovery (Keyes et al. 2021). Unlike algorithms for production, they do not have their output directly plugged into decisions like what price to quote or whether to offer medical treatment. Instead, they are used as research methods. Researchers reason about them to make knowledge claims, writing for example papers on what factors predict morbidity. This use necessitates considerable explicit attention to measurement, design, and theoretical implications of algorithms. Those implications are, of course, always present in algorithms, but the shift to discovery use from production use necessitates a shift on the part of algorithm designers/users from leaving them as unconscious assumptions to thinking through them as explicit theory. Thus it may be harder for algorithms to reproduce social inequalities in discovery contexts, because the primary mechanism for reproducing inequalities—inattention—is disincentivised in scientific research.

While scholars emphasize problems arising from *mis*alignment of theory, data, and measurement that occurs in production algorithms (Gebru et al. 2020; Jacobs 2021), others have shown that with the explicit attention of the discovery context, we can see better results. For example, noticing the alignment between machine learning methods and intersectionality theory allows for improved inquiry on specific research questions (Nelson 2021). Others argue ML "could offer powerful new insight" by tapping into new measurements—a "vast expanse of information about what people do, know, think, and feel" (Evans and Aceves 2016:22). Some even argue that machine learning will revolutionize scientific discovery in general by forcing us to use explicit, formal logic instead of the implicit biases of humans (King et al. 2009).

Some sociologists have remained skeptical, demonstrating that unexamined data, measurement, and theory assumptions can still cloud the use of algorithms for discovery (Gelman, Mattson, and Simpson 2018; Nelson 2019). Drawing on critical analysis of two papers using machine learning to study sexuality and autism, Keyes and colleagues argue that ML can serve "not to find the 'truth' of identity but to naturalize a particular view of it—one that, unsurprisingly, conforms with status quo assumptions" (2021:165). This is why Benjamin, who emphasizes the role of ignorance and inattention in algorithmic injustice, notes that racism "is not simply ignorance, or a not knowing. It is also ... a way of knowing the world" that we often do not recognize as such, causing us to "overlook it in the smart sounding logics of textbooks, policy statements, court rulings, science journals, and cutting edge technologies" (Benjamin 2016:148–49). Thus, I take up Seaver's call to "examine the logic that guides the hands, picking certain algorithms rather than others, choosing particular representations of data, and translating ideas into code" (2019:419).

### *2.1.3 Machine Learning in Neuroscience*

The rise of neuroimaging technologies like magnetic resonance imaging (MRI), gave scientists access to data sets with data sets that had between ten thousand and a million or more variables, each representing, for example, one cubic millimeter of a brain. This spurred the adoption of "mass univariate" analysis, in which single paper conducts many separate statistical tests, one to compare each variable to an outcome of interest (this is also how genome-wide association studies work). Scientists ask, "is this cubic millimeter of the brain correlated with sex? what about the one next to it? the next one?" and so on. While the mass univariate approach is often referred to as "atheoretic" because there is no specific theory motivating any of the individual statistical tests (Bzdok 2017; Haxby 2012), in truth it is premised on the usually-implicit "strong modularity hypothesis." That theory holds that "every possible face, animal, and object category has a specialized region [in the brain] or set of neurons dedicated to its representation" (Haxby 2012), an area that should 'light up' in brain scans when the right stimulus is applied.

But scientists do not believe the strong modularity hypothesis. Decades of failure with univariate methods lead neuroscientists and geneticists to turn to ever more complex, multivariate "witches' brew" and "psychobiosocial" explanations (Arribas-Ayllon, Bartlett, and Featherstone 2010; Eliot et al. 2021; Wade 2013). The strong modularity hypothesis "didn't seem possible. There are too many ways that faces and objects can be categorized" for each to have a dedicated part of the brain (Haxby 2012:2). Machine learning entered neuroimaging in early 2000s as an explicit attempt to match quantitative methods and models with new neuroscientific theory (Haxby 2012). ML tools were promising both because they may detect results where old methods failed (Lao et al. 2004), and because they have better alignment with theory (multivariate models for multivariate theories)—something algorithmic fairness scholars have been calling for (Jacobs and Wallach 2021).

But ML carries epistemic risks. Discussions of machine learning are often characterized by "enchanted determinism," magical discourses about the supposedly superhuman power of algorithms that are enhanced by "the inability to fully explain how these results are produced" (Campolo and Crawford 2020:1). Framing algorithms as advanced beyond human understanding also places them beyond human critique. This can lead to "machinic neoplatonism," or a belief that machine learning "reveal[s] a hidden mathematical order in the world that is superior to our direct experience," in which the "new symmetry of these orderings is more compelling than the actual results," results which often show the algorithms have substantial inaccuracies (McQuillan 2018:253). By placing such faith in algorithms, people risk the "thoughtlessness" that algorithmic fairness scholars say produces bad systems (ibid.). Thus it is not clear the extent to which algorithms for discovery will involve explicit reflection on the theory, methods, and measures that is typically missing from algorithms for production. Nor is it clear whether such attention would prevent algorithms for discovery from reproducing a harmful status quo.

### *2.1.4 Classification and Sex*

Classification plays a substantial role in social life and sociological analysis, because category systems are always moral and political (Geoffrey C Bowker and Star 1999:324), and "the categories that make [knowledge] possible are the stakes, par excellence, of political struggle" (Bourdieu 1985:729). Sociologists have long recognized for example that our assignment to a sex category has considerable impact on everyday life across social domains, so much so that it is even theorized as a 'master identity' (West and Zimmerman 1987). In their classic article, West and Zimmerman (1987) distinguish between sex, sex category, and gender in order to emphasize that placing people in a sex category is a routine and influential activity of everyday interaction, distinct from the biological criteria of their sex and the doing of gender. Its quotidian instances have substantial effects on our lives, especially when sex categorization 'goes wrong' (Keyes 2018; Lagos 2019).

The scientific use and production of categories, especially categories of people, has had considerable influence on society through shaping both individual and policy understandings (Igo 2007). As Benjamin notes, "social inequality is legitimized by cultural mythologies about human difference—stories that are made to matter through science, technology, and biomedicine" (2016:153). Scientific categorization plays a key role in 'making up people,' inventing new identities, ways of being and relating in the world. Hacking argues this process happens through several 'fundamental engines': "Statistical analysis of classes of people is a fundamental engine. Likewise we constantly try to medicalise... as early as the 1830s[,] the brains of suicides were dissected to find the hidden cause. More generally, we try to biologise, to recognise a biological foundation for the problems that beset some class of people," thereby creating or changing the meaning of that class (2007:293–94).

But scientific classification does not come as a view from nowhere. Instead, "the researcher often begins his [sic] classifications with only broad dichotomies, which he expects his data to 'fit,'... each classification level becomes a more refined measurement device for transforming common-sense meanings and implicit theoretical notions into acceptable 'evidence.'" (Cicourel 1964:21). Assigning bodies to sex categories is not just quotidian in the US today; it "was and continues to be a core technology of imperialist empire-building," inextricably linked with racializing bodies (Scheuerman et al. 2021:2).

### *2.1.5 The Case: Brain Sex Science*

Neuroscience research on sex is an ideal context to study the role of algorithms for discovery. Scientists' use of algorithms to make claims about the nature of sex (i.e. to argue for particular theories of sex) offers a poignant test case for theories of algorithms that were formulated by examining production algorithms. Existing theories emphasize that the point of algorithms is not the measurement or ontology of the social categories they operate on, leading to thoughtlessness and neglect which are diagnosed as the key problem. Here, the measurement and ontology of sex are the whole point of producing algorithms. How does that change our understanding of what algorithms 'do?' Moreover sex classification is important for society, and there is a long, rich tradition of research both doing and criticizing brain sex categorization that can contextualize new findings. Brain classification has a long history of social importance (Hacking 2007; Rollins 2021; Rose 2013), and sex has become the 'gold standard' for machine learning with neuroimaging data by which new algorithms and methods are evaluated (Yuan et al. 2018).

Considerable scientific effort has gone into producing and maintaining the sex binary, the notion that men and women are essentially, categorically distinct and that this distinction is 'natural' (Fujimura 2006; Laqueur 1990; Oudshoorn 2014; Sanz 2017). In particular, there is a long tradition of research on "brain sex" that seeks to explain all manner of personality, occupational, and educational differences between men and women as a function of innately different brains, thereby naturalizing and justifying gender inequalities (Eliot et al. 2021; Jordan-Young 2010; Lockhart 2020; Rippon 2019). Such work continues today, with a substantial scientific social movement behind it that has resulted in new journals, professional societies, and even a requirement by the National Institutes of Health that all funded research examine 'sex as a biological variable' (Epstein 2007; Pape 2021). The scale of scientific efforts to establish and defend the sex binary betray its highly contested status: the sex binary (a view holding that sex is two distinct, innate, biological, immutable categories) is far from settled science (Lockhart 2020; Richardson 2022; Sanz 2017). Indeed, sex research is plagued with replication failures, small effects, and discordant results (Eliot et al. 2021; Jordan-Young 2010; Richardson 2013). Sociologists of science note that sex scientists regularly encounter and ignore "awkward surpluses" of data that contradict their binary theory of sex (Fujimura 2006; Scheuerman et al. 2021).

None of this is to say that there is no material, biological substance to sex. Sociologists and feminist biologists have long argued for the importance of biology in sex (e.g. Fausto-Sterling 2005; Hird 2004a; Wade 2013). Their positions, however, represent an anti-essentialist, "feminist biology," "non-linear biology," or "sex contextualist" approach that stands in stark contrast with the essentialist "sex difference" approach (Hird 2004b; Lockhart 2021; Richardson 2022). In quantitative terms, there are three broad theories of sex in the scientific debate today. Those focused on sex differences posit a "truly categorical" difference in kind between male and female, which is aligned across biological and social traits (e.g. hormones, gonads, genetics, hair, immune response, muscle mass, aggression, verbal ability, toy preferences, occupation). They contrast this with a strawman "blank slate" theory that sex is a random vector, uncorrelated with anything. In truth, their scientific opponents argue that sex is the name humans give to the observation that a wide variety of things are correlated and mutually influential, but they stress that we understand sex best if we admit and investigate the variation within, among, and beyond sex traits and categories (Eliot et al. 2021; Keyes et al. 2021; Lockhart 2020; Miyagi, Guthman, and Sun 2021; Richardson 2022).

### 2.2 Data & Methods

In order to produce more complete accounts of technology and society, we must "follow social links even when they weave their way through non-social objects" such as algorithms and MRIs (Fitsch 2021; Latour 2005:83). Sociologists of science do so by examining the traces left when objects such as algorithms are made to account (unlike humans, objects do not speak on their own). Those traces are abundant during innovations and controversies in the documents produced in scientists' laboratories, those "that detail and lay out technical specifications [such as] conference papers on machine learning" (Bucher 2016:87; Latour 2005). More simply, if we wish to understand the social construction of scientific facts, we should examine how scientists construct facts: the data and methods they use, the interpretations they devise, and the rhetoric they communicate to peers.

Because this kind of analysis requires going into significant depth, I follow Mayrl and Wilson's Second Analytic Architecture and analyze one case (2020:1371). I selected a controversy in which algorithms and sex are being produced as both scientific and social objects, and I collected the artifacts of that production: the scientific reports; their comments, replies, and retractions; their supplemental information and code. I collected every publication that met the following criteria: written in English[4]; uses neuroimaging data (e.g. MRI); and uses modeling (broadly defined) to predict sex/gender. A recent review of brain sex differences research lists 12

[4] The vast majority of natural science papers are written in English, even in countries where authors' and audiences' first language is generally not English (Di Bitetti and Ferreras 2017). Authors of the papers in my sample are based around the world, including 22 countries on five continents.

studies using "multivariate statistical learning algorithms" to predict sex from brain imaging data (Eliot et al. 2021). I used those studies as a starting point, reading them for citations and keywords to search in both my university library's databases and Google Scholar. I checked each result for the inclusion criteria and saved the relevant papers. Throughout the process I added to and refined my set of search keywords, until the results were exhaustive (Schilt and Westbrook 2009).

During my initial open coding (Emerson, Fretz, and Shaw 2011), I reviewed each paper's citations and added four additional papers that I missed while searching. I also removed a few papers that turned out not to meet the inclusion criteria, including five studies measuring skull bones in order to identify human remains. The resulting data set consists of 53 papers (48 published and 5 preprints) through early 2021. For one software package used in a number of papers, libsvm, I accessed both the current code and historic versions of it back to 2012 on github in order to verify methodological details that were unspecified in the publications.

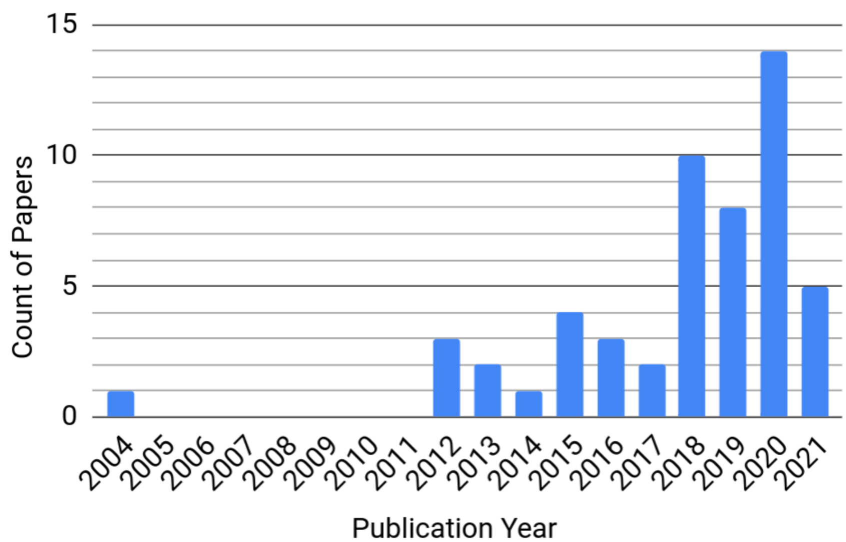


Figure 1: Papers using machine learning to predict sex from brain imaging data.

I draw on a mix of techniques from content analysis (Epstein 1996), actor-network theory (Latour 2005), and ethnography (Emerson et al. 2011) to analyze my data, a combination approach Bucher (2016) refers to as "technography." Technography observes technology to understand the interplay among people, with and through technology. It pays particular attention to how norms and values are delegated to and materialized in technology (Bucher 2016:86). After an initial inductive, open coding of the data, I revisited all of the papers using a focused coding approach for 23 factors that emerged during open coding (Emerson et al. 2011), including imaging technologies, model types, accuracy, sample size, validation strategy, feature selection method, number of features, and disciplinary audience. During this second pass, I also gathered and coded every reference to ethics, research motivations, etiology of sex differences, how sex was measured, substantive interpretations of the findings, use of the word 'discrimination', and references to critics of categorical brain sex difference research. These codes, along with my research memos, enabled me to develop insight by juxtaposing claims made in scientific literature (Waidzunas 2015:258).

### 2.3 Difference by Design

Most of the papers I collected fall into two main groups based on their stated goals. The largest group of papers, 26 out of 53, I call sexual dimorphism papers. These papers set out to show that human brains come in two categorically distinct forms, with titles like "Patterns in the human brain mosaic discriminate males from females" (Chekroud et al. 2016) and "Combined structural and resting-state functional MRI analysis of sexual dimorphism in the young adult human brain" (Wang et al. 2012). These papers are designed to reinforce our culture's dominant, categorical binary understanding of sex.

Seventeen (68%) of the sexual dimorphism papers include statements in their introduction asserting that sexual dimorphism (literally "two forms") in the human brain is already an established fact. In these, sex differences are both the explicit starting premise and the expected finding. Baldinger-Melich et al. (2020) make this their very first sentence: "Sex differentiation of all somatic tissues, including the brain, is driven both by direct genetic influences and gonadal hormones with only minimal environmental effects." Others assert that "The existence of organic differences in the brain, attributable to sex, has been well-replicated and reviewed at an aggregate level" (Anderson et al. 2019:1496) and that "Previous studies of brain structure in vivo using magnetic resonance imaging (MRI) have revealed consistent differences in whole brain tissue volume between the sexes" (Sepehrband et al. 2018:217). By finding that brains can be classified as male and female, ML does not challenge what the authors claim is established knowledge about sex and brains. It merely lends the imprimatur of fancy methods to the authors' explicit starting assumptions, thus perpetuating a centuries-long tradition of scientifically 'proving' men and women's brains are different (Laqueur 1990; Lockhart 2020; Sanz 2017).

This is especially evident in the response to Daphna Joel and colleagues' 2015 paper on the 'mosaic hypothesis,' one of many papers from critics of the sex differences approach that argue (without ML) that human brains are not categorically male or female. Instead, they suggest, individual brains are a mosaic of traits that are on average associated with each sex (Joel et al. 2015). Eight days after the paper was published, a team of scientists posted a preprint critiquing it. In order to argue that Joel is wrong and brains really are sexually dimorphic, they report using ML to predict sex from Joel's data, resulting in accuracies "ranged from 68.5% to 77.2%" (Del Giudice et al. 2015:8). Because ML could predict sex more accurately than random guessing (50%), they argued, the algorithm must be picking up on some real difference between men and women's brains. The preprint further suggests, and quantitatively tests, differentiating monkey species by face shape as an analogy to human brain gender classification. The preprint authors included polemic figures like Marco Del Giudice, Richard Lippa, and J. Michael Bailey, who have dedicated a substantial portion of their careers to establishing the sex binary and rebutting scientists who cast doubt on it (e.g. Del Giudice 2021; Del Giudice, Booth, and Irwing 2012; Lippa 2006). Within four months, *PNAS* published a shortened version of their preprint along with two other letters in response to Joel's article.

The other rebuttal letters use ML the same way. The author of one explains why he chose to use ML classification: "a classifier can only achieve perfect classification if the data points are well separated (note the converse does not hold: the data may be well separated, even if a particular classifier is no better than random guessing)" (Rosenblatt 2016:E6966). He believes ML can only help his cause of showing male and female are 'well separated.' He chose a method he believes can only criticize Joel's position, not support it. He then used the model to conclude, contra Joel, that "brains are indeed typically male or typically female" because ML can classify sex from brains more accurately than random guessing (2016:E6966). The third letter concedes "that a strict dichotomy between male/female brains does not exist." Nevertheless, its authors insist that human men and women's brains are indeed categorically distinct in the same way that cats and dogs are. Their ML results for classifying brains according to sex show "an individual's biological sex can be classified with extremely high accuracy by considering the brain mosaic as a whole," even though there is within category variation (which they analogize to "breeds of dogs") and no "singular physical characteristic reliably distinguishes cats from dogs" or men's from women's brains (Chekroud et al. 2016).

While some sexual dimorphism papers nod to uses for their results beyond simply proving the sex binary, the authors' revealed preferences tell another story. Claims about the utility of discovering sex differences for things like disease or personality research are often made only in passing, appearing in introductions and significance statements but not in the analysis. For example, Anderson et al. (2019) claim sex differences in brains are important for understanding criminal behavior, since men commit more violent crime. They use large data sets on both incarcerated and non-incarcerated populations. But the authors do not conduct analyses that could shed light on the relationship between gender and incarceration (much less crime). Instead, they narrowly "set out to examine the reliability of sexual dimorphism in the brain among a large incarcerated sample, extending work carried out by others in nonincarcerated healthy control subjects" (Anderson et al. 2019:1501). That is, they set out to prove that sexual dimorphism—which they claim in their introduction is universal and already demonstrated—also holds in another subpopulation. That reinforces the universality of the sex binary and tells us nothing about crime or what differentiates incarcerated and non-incarcerated people.[5] This is not accidental. Anderson and colleagues are aware of the debate about brain sex differences that they are stepping into. They spend most of their introduction summarizing it, placing their contribution squarely in the debate about the nature of sex rather than in criminology.

### 2.4 Sex as Gold Standard

The second largest group of papers (n=16, or 30%) are what I call proof of concept papers. These papers are interested in demonstrating a new method for analyzing neuroimaging data, and they use sex prediction as their example use case. Nevertheless, they function to produce scientific evidence for the dominant, binary conception of sex that they rely on. Unlike sexual dimorphism papers, three quarters which are published in neuroscience journals like *NeuroImage*, and the rest in general-interest journals like *PNAS*, proof of concept papers are divided evenly between computer science venues like IEEE conferences and neuroscience journals. Only one is published in a general interest venue, *Nature Communications.*

Proof of concept papers say almost nothing about sex/gender. For example, the paper "On the generalizability of resting-state fMRI machine learning classifiers" (Huf et al. 2014) is fundamentally disinterested in sex. The authors mention sex only parenthetically, saying

[5] Indeed, their findings tell us that incarcerated people are *not* different from others. This implication seems at odds with the tens of millions in research funding their lab receives to find the causes of crime in brains, and they do not make it explicit.

“classification accuracies of up to 0.8 (using sex as the target variable) could be achieved” with their proposed method (2014:1). Some proof of concept papers are clear about why they use sex to demonstrate their method: “we use the subjects’ gender label [male/female] as the predicting label because gender is the golden standard in the neuroimaging field and does not include subjective factors” (Yuan et al. 2018:49926). Others concur: “the classification problem is relatively easy, as sex can be unequivocally determined and brain sexual dimorphisms is [sic] well established” (Nieuwenhuis et al. 2017:248). Whether writing in computer science or neuroscience, they assume their audience does not need to be convinced that brain sexual dimorphism is an established scientific fact. Rather than framing the performance of their ML models as evidence for the scientific fact of brain sex dimorphism, the fact is taken as given and the performance of the models is interpreted as evidence of the models’ quality. In this way, they are more like computer scientists optimizing production algorithms than like other scientists using algorithms for discovery.

However, the publications from each group are all part of the same publishing and citation ecosystem. Proof of concept papers are published and cited contemporaneously with sexual dimorphism papers in overlapping journals. It is not that scientists first demonstrated that human brains come in two discrete categories (male and female) and then used these validated categories to test new methods. Instead, in a circular chain of logic, some scientists are using the accuracy of ML to debate whether human brains are sexually dimorphic at the same time that others are using sex dimorphism as a gold standard to test whether algorithms are accurate. Proof of concept papers, seemingly unaware of the controversy around brain sex dimorphism, nevertheless contribute to the accretion of evidence for it by adding to the list of publications ‘finding’ the dichotomy. By not acknowledging the controversy, such papers do more to shore up the sex binary than even the dimorphism papers, as the latter usually mention mixed results in their literature reviews and research limitations. Proof of concept papers are also generally aimed at different audiences. Sexual dimorphism papers are often written for general science audiences, to convince them of sex differences, while proof of concept papers are often written for computer scientists to share new methods for image classification. Even among papers in the same journal, such as *NeuroImage* or *Frontiers in Human Neuroscience*, the introductions, literature reviews, conclusions, and significance statements make it clear they have different audiences. Dimorphism papers are written for readers interested in the nature of sex in brains, while proof of concept papers are written for readers interested in the methods of analysis. This later audience, like computer science audiences, is less likely to be aware that brain sex classification is controversial among domain experts, and thus more likely to take claims about it at face value.

Yet even beyond proof of concept papers, the “unequivocal” and “not subjective” nature of sex is a design feature of nearly all the papers in my sample. Fourty-four say nothing at all about how sex was measured, while only five included biological measures (hormone levels, puberty, menopause, or chromosomes). These measures were included primarily to exclude intersex people from analysis (their bodies give the lie to the binary theory of sex). Three of the studies with such measures focus on trans people, leaving only two studies about cisgender people that measure any biological dimensions of sex. As Westbrook and Saperstein note, social science researchers tasked with collecting sex/gender data are “rarely given criteria to use in labeling” subjects’ sex, because essentialist beliefs present sex as too “obvious” for rigorous measurement (2015:545). I find that even for research on the neurobiology of sex, sex is generally taken as too obvious and binary to be worthy of measurement, exactly the process of inattention that algorithmic fairness scholars highlight in algorithms for production.

**2.5 High Hopes**

Machine learning methods were introduced into neuroimaging in the early 2000s explicitly as a solution to the strong modularity hypothesis becoming less tenable (Haxby 2012). The first paper to use them for sexing brains said so clearly in its abstract: “by applying multivariate pattern classification methods, [we] can detect subtle and spatially complex patterns of morphological group differences which are often not detectable by… [univariate] methods [that] analyze morphological measurements voxel-by-voxel and do not consider the entirety of the data simultaneously” (Lao et al. 2004:46). Twenty out of 26 papers published before 2019 have statements like this one, saying ML is promising because it will find sex differences where other methods have failed. By 2019, the use of ML in neuroimaging is more established and fewer papers make explicit comparisons to univariate approaches.

Some use diagrams with made-up data to explain this point (e.g. Huf et al. 2014; Lao et al. 2004; Wachinger et al. 2015). The most extreme example is Figure 2 from Rosenblatt (2016). Rosenblatt argues that even if no measures of brains appear categorically distinct in univariate analysis, men and women’s brains might still be categorically distinct in multivariate analysis.

He illustrates the point with made-up data showing an almost perfect separation between two groups that is only visible if we evaluate two variables at the same time. This is the promise of machine learning for authors seeking to prove sexual dimorphism. Despite using over 100 variables, Rosenblatt's ML model is unable to produce the clean categorical distinction his hypothetical diagram with two variables promises: it yields only 80% accuracy, indicating substantial overlap between "male" and "female" brains. Nevertheless, in a perfect example of McQuillan's (2018) "machinic neoplatonism," Rosenblatt concludes "given our empirical evidence and the *multivariate intuition depicted above*, we cannot help but disagree with [Joel] … Brains are indeed typically male or typically female" (Rosenblatt 2016:E1966 emphasis added). The theoretical symmetry of the algorithm, depicted in a figure with made up data, is compelling enough to override the actual results.

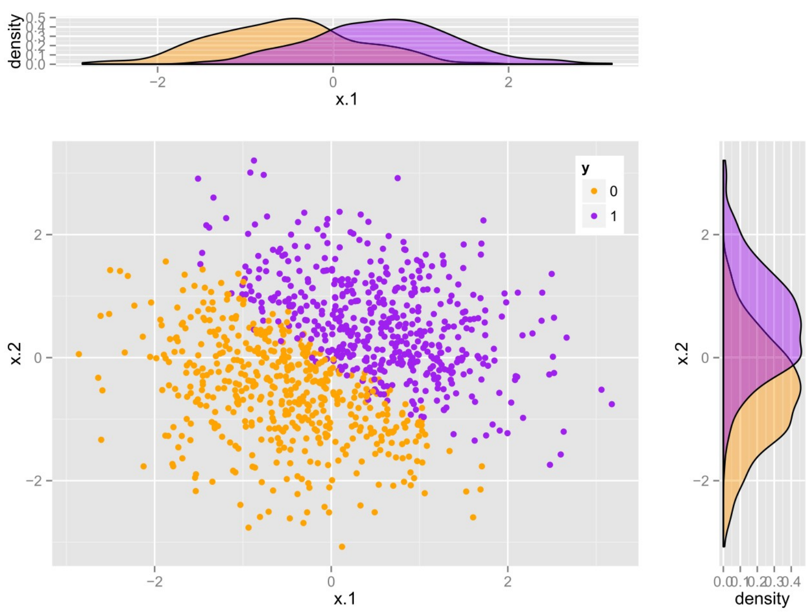


**Fig. 1. Univariate overlap with multivariate separation.**

Figure 2: Rosenblatt's (2016) hypothetical data justifying the use of machine learning.

Most papers claim ML does more than find previously undetectable differences between men and women. They promise to find a new kind of difference. Feis et al. write that "Crucially, because our approach is multivariate, it does not rely on individual voxels [variables representing areas of the brain] being discriminative by themselves. Rather, our approach enables us to detect combinations of voxels that are jointly informative about gender" (2013:256). Yet like others, the authors do not discuss "combinations of voxels" in their findings. Instead, they list out individual voxels (brain regions) in a series of tables and colorful brain diagrams. Readers are told that the dorsolateral prefrontal cortex, the superior frontal gyrus, and so on are each "regions discriminating between women and men" (254). There is no mention of relationships among these regions. Despite the methodological promise of seeking multivariate patterns, the authors extract and present univariate findings. Designing and presenting only lists of independently discriminative brain regions is a "value-laden choice" by scientists that "privileges achieving some kind of mathematical ideal over modeling the genuine complexity in the data" (Hancox-Li and Kumar 2021:5). A majority of papers in this field select only variables that are statistically significant in univariate tests either as the inputs to or as the reportable outputs of their ML models (Bzdok 2017). Each side of this apparent contradiction contributes to the papers' legitimacy. Multivariate theory and models are needed because univariate ones have become untennable. But univariate interpretation and reporting of those models remains the most intelligible approach to scientists trained in univarate analysis (Bzdok 2017), and 'locating' something in individual spots in the brain makes it seem more real and natural than it is (Roskies 2008).

### 2.6 Support Vector Machines and the Illusion of Accuracy

The frequentist statistical approaches that dominate neuroscience (and quantitative sociology) are about group properties that exist only in aggregate. Whether scientists are working with regression coefficients, two-sample t-tests, or ANOVAs, there is general understanding that the groups being compared are not categorically distinct, with their properties true of all members, but rather overlapping distributions whose properties are averages of diverse members. As a result, neuroscientists generally use "risk-thinking," focusing for example "on who *could be* violent rather than who *was* or *is* violent" (Rollins 2021:15 emphasis original). This puts neuroscientists trying to argue for categorical sex differences in a tricky place: to be accurate, they need to say that their findings are only true on average and do not apply to individuals. But to make their point that sex is *categorical* in the strong sense that they want to (recall that some of them make analogies to distinctions between animal species, or draw diagrams with nonoverlaping groups) using traditional statistics, they need to elide the messy overlap between groups and misrepresent the statistical 'average man' as universal (Igo 2007; Lockhart 2020).

Machine learning offers a way to escape this trap and speak confidently about categorical difference between groups. ML *is* about individual predictions and labels. Its focus is on 'y-hats' (predicting facts about individuals like sex or committing violent acts) rather than 'beta-hats' (risk factors, group averages) (Salganik 2017). Brennan, Wu, and Fan lay out the logic: "If the [ML model] can successfully classify biological sex significantly above chance using these features [brain scan data], this would demonstrate that dimorphism exists in the brain" (2021:2). This is fundamentally different from the frequentist approach: "while on-average differences in these regions have been recognized previously... it is conceptually important that the current work has demonstrated multivariate [ML] models that approach something closer to *truly dimorphic* patterns, effectively differentiating individuals into categorical groups based on intrinsic [brain] structural networks" (Anderson et al. 2019:1503 emphasis added).

To see how ML better fits theories of categorical difference, consider these descriptions of a particular ML algorithm, the support vector machine (SVM), from the brain sex literature:

> [SVM] is a classification algorithm that attempts to find a hyperplane which best separates binary classes (male and female) in hyperspace of predictive features (Brennan et al. 2021:2).
>
> the algorithm finds a hyperplane (i.e., a high dimensional plane) that maximizes the margin between the training samples of both classes. ... Kernel functions can be applied if the data are not linearly separable in the original space and allows for group classification based on non-linear effects (Anderson et al. 2019:1499)
>
> A hyperplane (or hypersurface) is determined that optimally separates the two groups of samples (Lao et al. 2004:49).
>
> linear SVM ... assigns a decision value to each subject, reflecting the distance between a given test subject's [brain] images and the hyperplane separating the two groups (Feis et al. 2013:254).

These are fairly standard descriptions of SVMs, echoing what one hears in machine learning classes and the primary SVM literature. They all imagine something like the image in Figure 2 and say that SVM will draw a line ("hyperplane" is the name for a line in data that has many variables instead of just two) perfectly separating the data points representing brains from men and women. Two of the quoted statements flatly assert that such a line is found, implying both that the task is possible and that the algorithm will succeed. One says SVM "attempts" to find such a line, but does not elaborate on what failure would look like. Another suggests "kernel functions" can be used to make SVM find a dividing line even when finding a line is not otherwise possible.

A method that can do what SVM promises has substantial allure for research on categories. If we can really draw a line separating two groups, as implied by these descriptions of SVM and the illustrations of made up data, then these scientists claim it would be strong evidence of those groups' binary, categorical nature. This tight alignment between theory and method may be why SVM is the most popular ML tool in brain sex research, used in 33 out of 53 papers.

Yet in practice, SVM's separating hyperplane is a fiction. The model generally does not converge on real data without a C parameter that allows some data points from each group to be on the 'wrong side' of the dividing line. The simplest evidence of this is the results in the papers themselves. None of the 53 papers presents 100% accuracy in predicting whether a brain is male or female. There are always some brains on the 'wrong side' of the line. The top reported accuracies from each paper range from 51% to 98%, averaging 82% and showing little improvement over time or as sample sizes increase. Indeed, others have argued that the lack of relationship between sample size and number of sex differences 'discovered' is evidence of reporting bias in this literature (David et al. 2018). Widespread design flaws in this research have also led to inflated accuracies, including failure to properly control for head size and other confounds (Dhamala et al. 2020; Eliot et al. 2021; Linn et al. 2016; Sanchis-Segura et al. 2020), failure to test on large enough and independent samples (Glocker et al. 2019; Huf et al. 2014; Varoquaux 2018; Zoubi et al. 2020), and failure to think through model assumptions (Bzdok 2017; Carlson et al. 2018; Schulz et al. 2020).

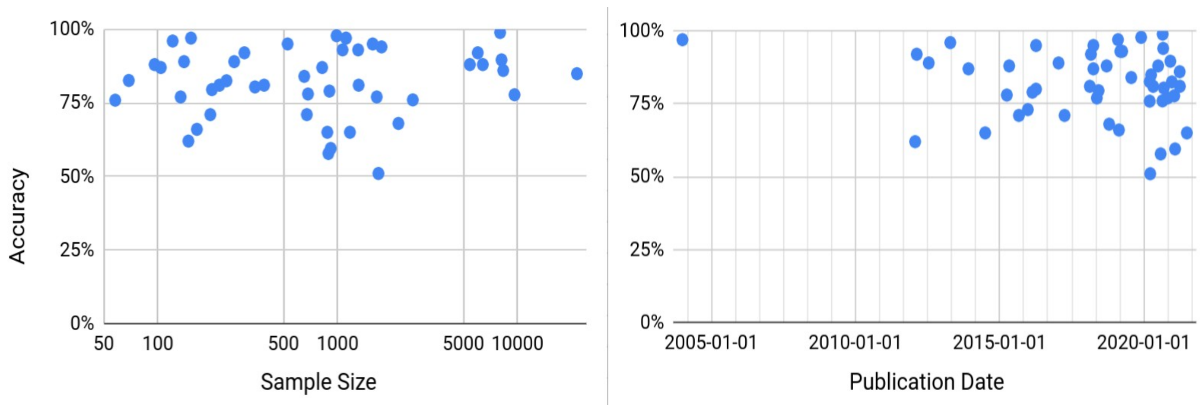


Figure 3: Accuracy of sex prediction from brain scans by sample size and publication date.

But even taking the reported accuracies at face value[6], how are accuracy scores mostly in the 70s and 80s presented as success in papers premised on the categorical separability of male and female brains? If a model 82% accurate[7], then it puts 18% of people into the 'wrong' sex category. The models fail to prove the sex binary by failing to separate male from female. Constructing the scientific fact of sexual dimorphism takes more work after the ML models.

To handle these disappointing results, sex researchers turn back to frequentist statistics. In a typical turn of phrase, Nieuwenhuis et al. tell us that "sex was predicted with significant accuracy (89%; $p < 0.001$)" (2017:246). Those familiar with ML might be puzzled: in ML one does not typically compute statistical significance.[8] Yet nearly every neuroscience paper using ML reports that their model achieved significant accuracy. They measure significance with permutation testing, in which scientists train the same ML model a thousand times, each with real brain scans (x data) but randomly assigned sex labels (y data). Then they compare the accuracy they get with the real sex labels to the distribution of accuracies they got on the random (fake) labels. This gives them a distribution for sex prediction under the "blank slate" null hypothesis that sex is a random vector uncorrelated with anything else. Because those advocating for sex differences and their scientific critics both agree that sex involves many things being correlated, it is not surprising that models with hundreds, thousands, or millions of variables find enough weak correlations to predict better than the random guessing strawman.[9]

[6] At least one of the papers retracted its findings after discovering a coding error (Ecker 2019).

[7] One gets 50% accuracy by tossing a coin.

[8] ML research typically prefers other measures like AUC, precision, and recall over accuracy as well.

[9] Only one paper in my data cites a source for this technique (Duarte-Carvajalino et al. 2012), and the paper it cites is an argument against using this significance testing approach in biomedical ML research (Ojala and Garriga 2010).

Significance testing ML models is used for more than discrediting the blank slate hypothesis, however. It is primarily used to argue that the binary categorical theory of sex is correct, even though the promised binary separation of groups did not materialize. The authors' theory of sex is embodied in these new categorical models. If the models performed as promised, drawing a line segregating men and women, the scientists would have their proof. When they do not, scientists are able to rescue the models, and thus the theory, by using significance testing. For example, Chekroud et al. (2016) report 93% accuracy before controlling for confounding variables, which reduce accuracy to 70%. Only then do the authors bring in significance testing, arguing that despite the reduced accuracy, the models "remained significant" after controlling for known confounders (2016:E1968). Van Putten, Olbrich, and Arns (2018) do something similar, initially setting up a high bar and then using significance testing to lower it: "We have excellent skills to extract sex from visual assessment of human faces, but assessing sex from human brain rhythms seems impossible," so they calculated "a significance threshold for the classification accuracy of 63%," meaning any accuracy above 63% is significant and sufficient to publish (2018:1). By turning from ML back to significance testing, scientists can assert that their categorical models of sex are correct despite the inaccuracies, because the models are "significant," which is often understood as shorthand for "true" or "correct."[10] The construction of scientific facts about sex takes work.

Each ML method used in this literature has its own narrative like SVM, and they share many of the same promises and challenges. Since 2018, neural networks / deep learning has become increasingly popular in this area, used in 14 papers (26%). Deep learning adds further credibility to brain sex research by promising to bypass the theoretical assumptions necessary for other types of models and operate directly on the "raw" brain image data to "learn the kernel" or functional form of the relationship between brain data and sex (Bzdok 2017; Carlson et al. 2018; Schulz et al. 2020:2). Such rhetoric is a form of 'enchanted determinism,' in which ML exceeds and replaces human capacity to understand patterns, further adding to its credibility and preventing critical reflection (Campolo and Crawford 2020). Yet "raw data is an oxymoron." This is especially true of neuroimaging data, which is always heavily processed (normalized, aligned, warped, motion-corrected, skullstripped, etc.) (Ichikawa 2021; Jackson et al. 2013;

[10] The more technically accurate, unstated interpretation is something like: "we would not expect to see results as good as these if the model was wrong." In machine learning, however, this intuition is reversed. If the model is poorly designed, we expect it to find very strong results (Lockhart and Weiss 2014).

Roskies 2008). So deep learning, like SVM, is used with an inaccurate, overconfident narrative in brain sex research.

### 2.7 Feature Importance

While the ML algorithms neuroscientists use are generally designed to optimize predictive performance at the expense of interpretable explanations, scientists nevertheless put considerable effort into 'decoding' how brains work using these models (Carlson et al. 2018). This decoding is valuable; 45 of the 53 papers enumerate lists of brain regions or measures responsible for the 'difference' between male and female. These lists are often illustrated with colorful brain diagrams and large tables. Such visuals increase the interest in and believability of neuroscientific claims, even among the neuro- and computer scientists conducting the research (Dumit 2014; Hancox-Li and Kumar 2021; McCabe and Castel 2008; Roskies 2008). As Figure 4 from Zhang et al. (2021) demonstrates: brain images are deployed even when they seem to lack face validity. The darkest areas on the right labeled "higher in male" are also the darkest areas on the left labeled "higher in female."

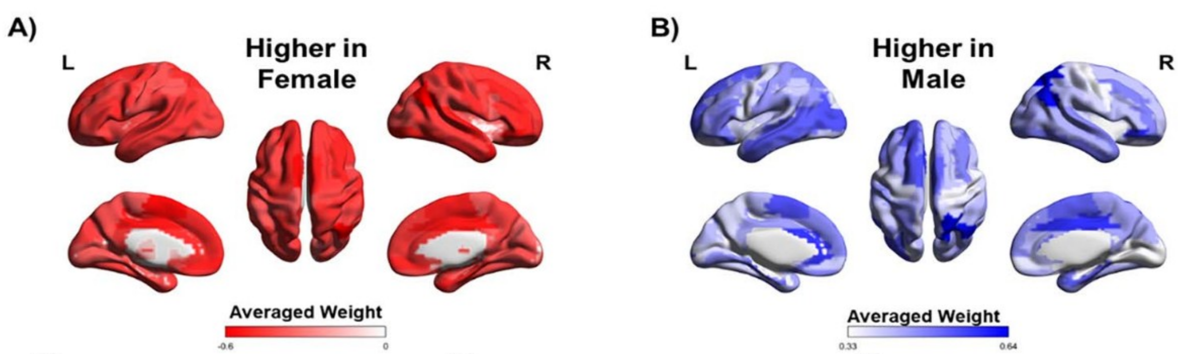


Figure 4: Image of brains from Zhang et al. (2021).

Identifying specific brain features also allows scientists to legitimate their findings by linking them to other research on sex differences. For example one team states, "we observed clusters of enlarged GM [gray matter] volume in the hypothalamus and both superior posterior lobes of the cerebellar hemispheres in boys... These clusters are centered on well-established regions of sexual dimorphism in the human brain" (Hoekzema et al. 2015:66). Here again, ML 'discovers' what the scientists claim to already know. Examined in detail, such apparently confirmatory findings in brain sex research are often contradictory, but the accretion of ever more studies with the same big-picture claims nevertheless lends the appearance of research consensus (Jordan-Young 2010).

Unfortunately, explaining the relationships ML models find and encode from data is difficult and often misleading (Hancox-Li and Kumar 2021; Rudin 2019). Unlike two-sample t-tests or OLS regression, where we can discern the size and direction of a relationship or difference in a fairly straightforward way (is the number big or small? positive or negative?), most ML models are more opaque. Let us return to the most popular model in this literature. SVM outputs weights (coefficients) for each input variable that are often interpreted as the "feature importance," that is, telling us how valuable those variables (called features) are for dividing brains into the categories male and female. As with other methods (such as ordinary least squares regression), a change to a variable with higher weight has more influence on the model's predictions than the same change to variables with lower weights.

But the meanings of SVM weights are difficult to interpret. A weight may be large because that variable is larger on average in men or women, or because that variable has a relationship with some other variable. Which of these cases is happening, and what other variable(s) the relationship involves are not discernible from the weights (Feis et al. 2013). Further, SVM often involves the use of nonlinear kernels, which transform the data using polynomial, radial, Gaussian, or sigmoidal functions that make relationships much more complex to interpret.[11] It was also common practice until the last few years to use dimensionality reduction tools such as principal components analysis on the data before ML, further transforming the data away from interpretability.

As a result, a few of the papers I study conclude "complicated [machine] learning frameworks make it difficult to infer the biological significance underlying the estimations" (Chen, Cao, and Tian 2019:2). Indeed, of the 45 papers that identify important brain measures, eight use only a separate univariate analysis to do so, noting that the ML models are too difficult to interpret. In these cases, the ML models serve only to prove the fact of difference, not to offer further insight about the nature of difference. Most authors, however, attempt to interpret ML weights. For example, although Feis et al (2013) explain why SVM weights have opaque meanings, they nevertheless dedicate nearly three full pages of a seven page manuscript to tables

[11] Despite this, a number of papers using SVM and interpreting the feature weights did not specify which kernel they used in the main text or supplemental information. The most popular tool for implementing SVM in this literature, libsvm (Lin and Chang 2021), uses a radial kernel when the user does not specify one, a fact I found in the source code on github, but not in the software documentation for users.

and figures interpreting their SVM weights, including 18 colorful brain images (red for female and blue for male, naturally).

SVM is not the only ML model used in this literature, but these difficulties of interpretation are widespread. Neural networks are a particularly illustrative case. Neural network research has focused heavily on image analysis and therefore provides a wide variety of tools for picking out what features in an image (like a brain scan) are related to a classification (like 'male' or 'female'). Most of these tools are forms of saliency mapping, which presents a heatmap of areas in an image that are "important" for classification. While the methods for generating saliency maps are complicated and diverse, the result is a list of weights, one for each pixel / voxel / point in the image that indicates how much a change in that variable would influence the model's prediction. In this sense, saliency maps are analogous to SVM weights or OLS coefficients. Like weights in SVM, those in neural networks cannot be interpreted in the simple, additive linear way that OLS weights can be. Usefully, because the weights are tied to variables that each represent points in an image, saliency maps output color-coded pictures of brains showing what parts are "important" for sex classification.

But neuroscience has a crucial difference from traditional uses of saliency maps like labeling whether a photograph contains a cat. In those uses, humans can generally look at the image and know whether and where there is a cat. This ability for humans to verify model output is central in the image algorithm bias literature (Schwemmer et al. 2020). Therefore, saliency mapping in computer vision applications is primarily used to evaluate whether models are good. There are numerous canonical examples of saliency maps revealing errors in models. In one, a neural network that could accurately classify pictures of animals as either Siberian huskies or wolves turned out to have saliency maps that showed the most important parts of the image for the model were not the parts with the animals, but rather their surroundings. Specifically, areas with snow were highlighted as evidence that the image was a wolf (Ribeiro, Singh, and Guestrin 2016). The algorithm did not learn differences between wolves and huskies. It learned that wolf pictures have snowy backgrounds. Others have demonstrated by human inspection that saliency maps can often be unreliable (Adebayo et al. 2018; Bansal, Agarwal, and Nguyen 2020). But humans cannot look at a brain scan and know whether the subject is a man or a woman. The point of modeling here, and with algorithms for discovery generally, is that we do not know what "maleness" or "femaleness" looks like in brain images, so we cannot check whether the model is pointing us to something real or to the equivalent of snow behind a wolf. Rather than using saliency maps as tools for diagnostics on models that potentially misrepresent sex, scientists accept the saliency maps as accurate representations of sex in the brain (e.g. Adeli et al. 2020; Arslan et al. 2018; Brennan et al. 2021; Górriz et al. 2018).

### 2.8 Another Way

Machine learning does not inherently lead to the reification of harmful cultural notions about binary sex, however. I found four papers demonstrating alternatives to the sex differences research program, using machine learning to advance positions more in line with feminist biology.

As Zhang at al. (2021) demonstrate in their paper, "The Human Brain is Best Described as Being on a Female/Male Continuum," scientists can interpret the same ML classification methods in a way that contests the categorical understanding of sex. The authors take advantage of the fact that SVM, like other ML classification methods and logistic regression, outputs predicted probabilities (e.g. "74% probability this brain image came from a man"). Typically researchers apply a threshold and say anything over, for example, 50% probability is male and anything below that is female, thus turning a continuous measure into a categorical one. Zhang et al. skip this step. Building on Joel's mosaic brain theory (2015), they assume all brains are a mix of "male" and "female features," then use the probabilities from their SVM to place brains on a spectrum rather than in binary categories. They show that the predicted probability of being male in brains is correlated with self-reports of gendered behavior, beliefs, and psychological metrics. Zhang et al. conclude that "The moderate classification accuracy [78%] of the multivariate classifier indicated that the brain functional architecture was unlikely to be conceptualized as binary, as is the case with biological sex, but was more likely to be continuously represented on a brain gender spectrum" (2021:11). Even more remarkably for this literature, they argue that "androgyny" in brains and behavior "is advantageous for mental health" (2021:11). In other words, the sex binary is not just scientifically inaccurate, but actively harmful to people. Using the same tools as many other papers (fMRI, SVM) with the same results (78% accuracy), the authors come to a radically different conclusion.

Sanchis-Segura and colleagues (2020, 2021) take a different approach. They still use the same kind of data (structural MRI) and ML classification methods (SVM, neural networks, and

10 others). But rather than building and trusting a single model, or building multiple models and reporting the results from the best performing one, they construct a variety of models with different measurement assumptions, then compare the results to understand how assumptions about sex influence the findings of neuroscience. They demonstrate that whether and how researchers correct for head size (total intracranial volume, TIV) and how they operationalize sex categories both have a substantial impact on the accuracy of ML models. The most reliable way of controlling for TIV reduces the average accuracy of models to 57%, and allowing an indeterminate sex category reduces accuracy to 43%.

Choices like how to handle head size depend on researchers' assumptions. Sanchis-Segura et al. (2020:12953) point out that others have observed large discrepancies between accuracies using corrected and uncorrected data, but chose to emphasize the uncorrected results. So what one believes about the importance of head size (or, cynically, what results one wants to find) is a critical aspect in choice of methods to use and emphasize. Studies feeding 'raw' brain scan data into ML models are not as atheoretical as they appear; they are implicitly taking the contested stance that raw size matters. By comparing models, Sanchis-Segura's team is able to avoid Rosenblatt's (2016) trap that classification models can only support claims of categorical difference. Their comparative approach demonstrates the role of researcher assumptions in findings of difference, helping to de-naturalize the sense of objectivity and undo the enchanted determinism around findings of sex difference.

In the most innovative paper, Joel et al. (2018) look beyond classification methods to other tools from machine learning: anomaly detection and clustering. Anomaly detection is an intuitive approach to the categorical difference question, because it tests whether "brain type(s) typical of females [are] also typical of males" (Joel et al. 2018:5). These models are given training data for something (e.g., brain scans from women) and then asked to predict whether new data (e.g., brain scans from men) appear typical of the data they have already seen, or anomalous. If human brains really come in categorically distinct, sexually dimorphic types in the way that the faces of different species of monkeys do (Del Giudice et al. 2015 made this analogy in response to Joel's 2015 paper), then 'male brains' should fall outside the norm for 'female brains' and get flagged by anomaly detection much more often than 'female brains' do. They do not.

Joel et al. (2018) also use clustering algorithms (which include tools like topic modeling). The logic here is straightforward: if brains really come in distinct forms according to sex categories, then tools designed to cluster the data into categorical groups should return groups that are largely sex-segregated. Unlike classification methods, where the computer is tasked with 'discriminating male and female brains,' clustering methods are tasked with finding the best division of brain data into groups without considering sex. When Joel's team compare the groups produced by the model with participant sex, they find poor correspondence. By using ML in which categorical sex difference is not baked into the design as an outcome variable, Joel's team is able to imagine and find alternate ways of understanding sex and variation in brains.

Unlike the sexual dimorphism and proof of concept papers, all four of these papers focus their introductions on reasons to doubt the sex binary. While introductions to academic papers are often written after the fact to match the findings, the prior work of these authors demonstrates a sustained commitment to the feminist biology tradition. This highlights an important factor in whether a given paper will retrench or challenge the sex binary that trumps factors like data and methods, as well as some proposed solutions like such as close attention to theory and measurement. It matters whether the authors want to challenge or defend the sex binary.

**2.9 Conclusions**

The use of machine learning in research on sex and brains is mostly a bait and switch: readers are promised more complexity and novel insights, but instead find the same old essentialism. Sustained criticism on both cultural and biological grounds, along with widespread replication failures, threaten the status of dominant, binary, essentialist approaches to sex research (Eliot et al. 2021; Lockhart 2020; Sanz 2017). In response, some sex scientists have abandoned traditional univariate modeling approaches and reductionist theories in favor of multivariate analysis and increased complexity—methodologically embodied as machine learning. ML is purported to solve both social and scientific problems, offering a path to a "fully post-phrenological neuroscience" (Anderson 2014). Yet in sex research, ML studies are often designed and interpreted so that their effect is to make sex appear even more binary and essential than traditional methods.

Brain sex research uses algorithms for discovery to retrench the sex binary and resist social change in multiple ways. Sometimes this happens through inattention—the primary mechanism emphasized in both algorithmic fairness and gender scholarship. For example, brain sex studies often consider sex—their primary outcome variable—too obvious to merit rigorous measurement (Jacobs and Wallach 2021; Westbrook and Saperstein 2015). This work also uses the rhetoric of enchanted determinism, saying ML transcends the limits of older methods that are grounded in conscious human theorizing about the data, thereby positioning the findings as beyond human critique (Campolo and Crawford 2020). Proof of concept papers may be the ultimate 'thoughtless' reification of the sex binary and machinic neoplatonism (McQuillan 2018), since an entire paper about a sex classification algorithm might only mention sex one time, in parentheses.

Importantly however, algorithms for discovery also retrench the sex binary through conscious reflection. The authors are often explicit about setting out to show sexual dimorphism or discrediting more complex, feminist and nonbinary understandings of sex. Scientists consciously shifted their methodology from univariate approaches that largely failed to show sexual dimorphism to supervised classification algorithms, which are much better aligned with their theory of sex as "truly dimorphic" (Anderson et al. 2019). This is an important rejoinder to algorithmic fairness scholars' argument that algorithms reproduce the problematic status quo through unnoticed mismatches between theory and modeling (Jacobs 2021). In brain sex research, conscious attention resulted in tight alignment between theory and model, but did not prevent algorithms from reproducing old, harmful social structures of sex essentialism.

The properties of ML alone do not cause these effects. Using methods designed to find a dividing line between male and female does not result in actually finding such a line. Instead, scientists must do extra work to produce a binary by, for example, using significance testing to argue that their categorical model is correct, despite its failure to produce distinct categories. Sex difference scientists must also ignore the warnings in both their references and their own text in order to conclude that their list of important brain measures has an interpretable meaning. As the feminist biology papers demonstrate, algorithms for discovery can just as well be used to deconstruct the sex binary and advance a more complex understanding of sex in human brains. Researchers can interpret the same ML tools differently, use them differently, or use different ML tools in order to achieve different effects.

My findings point to an important omission in theoretical work on the role of machine learning and algorithms in society. While a wide variety of work demonstrates that algorithmic systems often end up retrenching old social biases, their explanations turn on inattention and thoughtlessness. The same is true for critiques of science (Cicourel 1964; Westbrook and Saperstein 2015). Predictive policing or automated resume screening will have racist and sexist outcomes because they are trained on data that reflects longstanding racist and sexist institutions. Racism influences who works in tech, and what faces image labeling systems are developed on and for, thereby influencing who these systems recognize as human or as existing at all (Benjamin 2019). This inertia can be so great that it defies algorithm designers explicit efforts to remove bias (Schwemmer et al. 2020). In neuroimaging research on sex, a parallel finding would be that our heavily gendered society influences how brains develop, leading to patterns of difference that ML systems pick up on and then perpetuate as 'natural' and 'objective.' Undoubtedly, this happens. Society shapes the input data (brains) (Rippon 2019; Wade 2013), and I document that we observe biased output from the AI models of brains. However, this is not my argument.

Algorithms for discovery operate to retrench social biases in a different way than previous work theorizes. ML is deployed with the express intent or assumption of binary sex essentialism, and through a variety of design and interpretation choices that I unpack, scientists use ML to conclude that binary sex essentialism is the best understanding of the world. This finding stands in stark contrast to the promises of radical social change made by many studies of algorithms. Social scientific work on algorithms has concluded that they are "used to produce new versions of the world—versions that might differ greatly from their nonalgorithmic counterparts" (Cheney-Lippold 2017:33) or have less bias (King et al. 2009). Algorithms may be transforming "almost every major social institution" (Burrell and Fourcade 2021:1), but algorithms are not only or necessarily a tool for change (good or bad). They are also sometimes tools for *preventing* social transformation, for blocking new versions of the world, for telling us that a categorical distinction between men and women rooted In biology cannot be contested.

Brain sexing is not the only use of algorithms for discovery. Neuroimaging and ML are also used to study age, crime, language, emotion, intelligence, autism, depression, and much more. ML is increasingly used for discovery in genomics research on race/ancestry, intelligence, sexuality, personality, and myriad other important dimensions of social stratification. Computational social science is increasingly adopting machine learning to study human groups and categories as well, in ways that can be good or phrenological (Gelman et al. 2018; Kosinski

2021). As machine learning is taken up in a quest to answer more and more research questions, it is increasingly important for us to ask how the social construction of scientific facts and kinds of people proceeds in these cases, rather than relying only on our existing theories of algorithms, which were developed in a fundamentally different context, their use in production.

## Chapter 3 Presenting the Work: Historical Authority in the Science of Sex[12]

> Scientists' testimonies are used to endorse everything from toothpaste to nuclear power and weapons, but they are also used to challenge the very same things. And this is where the knife goes in because at present "scientific" support can be elicited on all sides of every question, so the "lay" public is constantly forced to decide which scientists to believe.
>
> Where then is the vaunted objectivity of science? People are realizing that they must… develop criteria on which to make these decisions (Hubbard 1990:9).

Notions of essential, biological sex differences play a major role in contemporary social and policy debates, ranging across the under-representation of women in science, government, and corporate leadership; the division of household labor and childcare; the access and rights of trans and intersex people to use appropriate facilities or to exist at all; and the best way to educate boys and girls. Far right movements have taken up essentialist arguments about biological sex in service of their agendas (Ferber and Butler 2020; Mears 2020; Stacy 2020). These arguments by the right rest their legitimacy on the authority of science and sex difference research. Many have engaged the substance of sex difference research shown how essentialist conclusions about sex are unwarranted in the science (Eliot et al. 2021; Fehr 2020; Richardson 2022). In this chapter, I take a different approach and use sociology of science to examine the competing claims to authority made by scientists studying sex. I argue that historical revisionism is a key means of establishing authority for scientists who advocate "essential sex differences," and that this undermines the credibility of their claims.

More than almost any other field of scientific research, sex difference scholars push their findings to general audiences. A quick search for books with "sex difference" in the title returns more than 2,000 volumes, in addition to the torrent of interviews and op-eds on the topic that

[12] A version of this chapter has been previously published as Lockhart, Jeffrey W. 2020. "'A Large and Long Standing Body': Historical Authority in the Science of Sex." In *Far Right Revisionism and the End of History: Alt/Histories*, edited by Louie Dean Valencia-García, 359–86. New York: Routledge.

researchers give to the popular press.[13] Statements like this one in the *Los Angeles Times* are routine: "the scientific reality is that it's futile to treat children as blank slates with no predetermined characteristics. Biology matters. A large and long-standing body of research literature shows that toy preferences, for example, are innate."[14] Despite what proponents of essential differences would have us believe, there is also a large and longstanding body of research literature that is critical of the "sex difference" paradigm. Many scientists have challenged the scientific basis for claims of essential sex differences, arguing that biology is more complex, less deterministic, and less suited to categorical binaries than sex difference scholars claim. They include Ruth Bleier, Katherine L. Bryant, Gillian Einstein, Lise Eliot, Anne Fausto-Sterling, Tristan Fehr, Cordelia Fine, Geordana Grossi, Donna Haraway, Ginger Hoffman, Ruth Hubbard, Janet Hyde, Daphna Joel, Rebecca Jordan-Young, Anelis Kaiser, Marion Namenworth, Gina Rippon, Joan Roughgarden, Deboleena Roy, Rafaella Rumiati, Sigrid Schmitz, Stephanie Shields, Abigail Stewart, Banu Subramaniam, Sari van Anders, and Mariamne Whatley. Far from being anti-science, these scholars have dedicated much of their careers to biological research.

I call these researchers and their work "feminist science," a term many scholars who challenge essentialist "sex difference" research have taken up (e.g. Bleier 1986; Roy 2012). Challenging the sex difference paradigm in biology does not mean insisting that men and women are identical. Instead, feminist biologists emphasize three points. First, biological traits show much more variation within groups (like men and women) than between them, which makes speaking categorically about things like male and female brains nonsensical. For instance, although the average height of men is greater than the average height of women, knowing a person's height tells us little about their sex, and vice versa. Indeed, while the field of sex differences relies on arbitrary cutoffs to define "small," "moderate," and "large" differences (measured as Cohen's $d > 0.2$, 0.5, and 0.8), these are all dramatically smaller than the sex difference in height ($d = 2.0$), which statisticians point out is still *not* bimodal (Schilling, Watkins, and Watkins 2002). This means that when we look at height data for all humans together, we do not see two separate groups in need of explanation by way of sex, but rather one group (one bell shaped curve) with all sexes mixed throughout. Second, many biological traits

[13] This holds in most large catalogs, and is discussed at length in (Bluhm 2012).

[14] See also (Gur and Gur 2017); Versions of this debate go back to John Locke's "tabula rasa," and more recently Steven Pinker's critiques of "The Blank Slate."; (Soh 2017).

result from or get modified by social and environmental experience. Everything from testosterone levels and brain morphology to even height and menstrual cycle has been shown to vary dramatically depending on experience and environment. In other words, biology is not immutable destiny. Third, speaking of "sex differences" as essential results of innate biology often serves to reify harmful stereotypes and resist progressive calls for social change: *why fight against human nature?*

The terms of debate around sex differences are fraught. Neither feminist biologists nor sex difference scholars are homogeneous groups. Many resist simple labeling. For example, interdisciplinary feminist scientist Rebecca Jordan-Young at times separates the substance of her scientific analysis from her feminist convictions in order to address other scientists on their own terms (2010:9, 200). Other feminist scientists integrate the two more often in their argumentation style, but neither of these approaches is necessarily more feminist or more scientific. Simultaneously, many proponents of the sex difference paradigm refer to themselves as feminists or liberals, including prominent figures like Simon Baron-Cohen and Melissa Hines. Most proponents of sex difference research distance themselves from "conservatives" or "the right." Even Charles Murray, writing for the American Enterprise Institute, tries to distance the science of essential, biological sex and race differences from conservative politics (2009). Steven Pinker argues that sex difference research and even *The Bell Curve* are "liberal," despite the authors' and their fans' right-wing politics (2002). Such rhetorical moves are typical of attempts to reconstitute neoliberal right-wing positions as apolitical or a "rational center" (Duggan 2003). As a result, the very terminology and "sides" in this debate are contested.

How are we to decide between the competing claims of these scientists? Their claims to authority often come down to competing historical narratives, either explicit or implicit, about the nature of sex difference research to date. In this chapter, I document three common types of historical revisionism used to bolster the authority of claims about sex differences. We have already seen the first type. The *Los Angeles Times* op-ed quoted earlier asserts a history in which innate, biological causes of social differences have long reigned as uncontested scientific facts. Such a history is blatantly revisionist, and the feminist biologists who are written out of that narrative typically respond by presenting extensive histories of debates within sex research to show that claims of innate, categorical differences are not uncontested. More subtly, many sex-difference publications present revisionist histories through citational practice, selectively citing only supportive material or even placing references next to ideas that they do not support (Young and Balaban 2006).

In a second approach, sex difference scholars position themselves as historical underdogs, defenders of Science, Truth, and Free Inquiry against the tyranny of "political correctness," trans activists, and feminists (Roy 2012:189; Young and Balaban 2006:634). This version of history is at odds with the first, in which sex difference reigns unchallenged. More to the point, I show it is not historically defensible either. Finally, the third type of revisionism involves setting up and burning a straw man I call "the big, bad social constructionist." By selectively reporting on and demonizing their critics, sex difference scholars are able to avoid substantive engagement with alternative explanations for the gendered world we live in. While the details of citations can sometimes feel peripheral to the main point of science, they can have far-reaching consequences for perpetuating baseless "academic urban legends" that translate into public policy and popular belief (Rekdal 2014).

In the rest of this chapter, I explore these claims in more detail. First, I sketch a brief history of sex difference research in order to clarify its political origins and rhetorical tactics. Then I present and evaluate three revisionist narratives common to sex difference research. Finally, I conclude by making explicit the role of scientists' motives in modern sex scholarship and calling for deeper engagement by both scientists and the public.

### 3.1 Eternal Return

Before engaging with revisionist accounts, it is helpful to review some often-omitted aspects of the history of sex difference research. While writing this section, I entertained a once forgotten fantasy: to write something using a patchwork of quoted material, with no original words of my own. One certainly could. Numerous books and articles have been dedicated to the critical history of sex difference research (Klein 1946; Laqueur 1990; Reinharz 1986 to name a few; Sanz 2017). Beth Hess summed it up well when she wrote,

> For two millennia, "impartial experts" have given us such trenchant insights as the fact that women lack sufficient heat to boil the blood and purify the soul, that their heads are too small, their wombs too big, their hormones too debilitating, that they think with their hearts or the wrong side of the brain. The list is never-ending (1990:81).

Critiques of biological sex essentialism are well established (e.g. Hacker 1953; Lowie and Hollingworth 1916; Poulian de La Barre 1673). Three decades ago, feminist biologists lamented the ongoing need to be "going over … old ground," and today they are still writing critical responses to "Whac-a-Mole Myths" of sex difference research (Fausto-Sterling 1992b:259; Rippon 2019).

But it does not take a kitchen table covered in feminist biology and history of science to catch on to the main thread of these arguments. They all show cases where the science of sex differences shifts over time in response to social beliefs and scientific advancement. Time and again, the social advancement of women motivates new waves of research on women's "essential character," from the suffrage movement, to the feminism of the 60s and 70s, to the moment when women began outperforming men in education. Each time, sex difference scientists are explicit that they are reacting to feminist movements (Fine 2010; Hubbard 1990). The purported biological basis of sex differences changes as science advances: women's brain fibers were prone to snapping until we discovered they were not; then their brains were too small, until we discovered brain weight does not correlate with intelligence (Fine 2010:xxiv). The frontal and then parietal lobes were each, in turn, too small in women when those areas were seen as the locus of intelligence (Shields 1980). Then women's corpus callosum was different from men's, until we found that it was not (Fausto-Sterling 2000:5). And now when feminists point out that the corpus callosum research was unreliable, advocates of sex difference say "of course this is completely unfair," because the latest sex difference research has shifted yet again to new measures of the brain (Cahill 2014:577). Whac-a-Mole indeed.

In order to make the political stakes and internal logic of sex differences clear, I trace out two threads of its intellectual lineage in more detail. The first highlights the political motives of this science and begins at least with Thomas Hobbes, whose seventeenth century commentary on the state of nature described it as a war of all against all, intense competition for survival and dominance. A century later, T. R. Malthus published his *Essay on Population,* which describes human races competing to the point of "extermination" through reproduction and argues that poor children should be left to starve as natural/divine punishment for their parents' choice to have children ([1798] 1958). Enter Charles Darwin, who read Malthus and often credited him as inspiration for his work on evolution.[15] Contemporaries Marx and Engels pointed out that

[15] Darwin is not alone in his Malthusian origins: like many sociologists and demographers, I was employed at a Population Studies Center as I wrote this.

Darwin's theories read directly as a transposition of Hobbes, Malthus, and nineteenth century liberalism onto the "natural" (non-human) world, complete with markets, competition, and specialization of labor.

In turn, Herbert Spencer and others brought these ideas back into the human world as Social Darwinism. In the end, "Darwin consciously borrowed from social theorists such as Malthus and Spencer some of the basic concepts of evolutionary theory. Spencer and others promptly used Darwinism to reinforce these very social theories and in the process bestowed upon them the force of natural law" (Hubbard 1990:90–92). Human social proclamations are not a perversion of some pure, objective, nature-focused Darwin. Darwin himself infamously wrote in *The Descent of Man* that "the chief distinction in the intellectual powers of the two sexes is shown by man's attaining higher eminence, in whatever he takes up, than can woman" ([1871] 1901:576). Evolutionary theories began as social theories of human aggression and hierarchy, and they have remained social theories.

While scientists rarely claim the term "eugenics" for their own work after 1970, eugenics research has an "openly continuous history" to the present day (Bashford and Levine 2010:542). In 1969, the *Eugenics Review* renamed itself the *Journal of Biosocial Science*. That same year, *Eugenics Quarterly* changed its name to *Social Biology*. Both continue to publish in 2022. In this same vein, E. O. Wilson published *Sociobiology,* a hugely influential work that sparked what is now known as evolutionary psychology ([1975] 2000). Wilson and evolutionary psychologists frequently cite Darwin's theory of sexual selection to argue *a priori* that there must be innate, cognitive differences between human men and women, just as there are differences between the tails of male and female peacocks, because of how evolution works (For example: Cahill 2010; Geary 2010). They assume that everything, including complex social processes and historically recent behaviors, necessarily serves some evolutionary purpose. Whatever we do today, they argue, must have been advantageous in the distant evolutionary past, when it was "hardwired" into our genetics for future generations. And so we get claims that men are good with maps and spatial reasoning because prehistoric men went out hunting, and that female monkeys prefer "feminine" toys like cooking pots while males prefer "masculine" police cars (Alexander and Hines 2002; Pinker 2002). These positions represent a staunch opposition to social change: social life today is the necessary *telos* of millions of years of evolution. Social change is therefore against human nature. Such claims have been met with substantial empirical and theoretical criticism from social scientists and biologists alike (Hubbard 1982; see also a related critique in Panofsky 2014; Quadagno 1979; e.g. Roughgarden 2004; Shields 2016).

The second intellectual lineage illustrates another core idea from feminist biology: the construction of categorical difference from complex phenomena. The history of "sex hormones" has been documented extensively (Fausto-Sterling 2000:6; Hall 1976; Oudshoorn 2003; Vines 1994). Gonads, especially testes, have been considered the source or essence of gender in many historical periods. In 1889, Charles-Édouard Brown-Séquard published the results of injecting himself with crushed guinea pig and dog testicles, claiming to experience increased virility and youthfulness. Although he admitted within a decade that the results were likely a placebo effect, the scientific quest to find the chemical essence of sex was on. By the 1920s and 1930s, substances called "male hormone" and "female hormone" had been isolated from gonadal tissue.

Almost immediately however, it became apparent that males – even stallions! – also had "female hormone" in their bodies, and vice versa. Scientific advances showed that testosterone and estrogen are part of a larger family of steroid hormones; that they get converted into one another within the body; that they are also produced outside gonads; and that testosterone and estrogen are both necessary for the regular functioning of non-sexual parts of the body, such as blood. In other words, the substances called "male and female hormone" were not as categorically distinct in form, function, or distribution, as scientists initially thought. Nevertheless, their association as categorical "sex hormones" remains today, defining the "true" essence of sex in professional athletics and justifying all sorts of gendered behavior (Davis 2015; Fausto-Sterling 2000:1–4; Fine 2010; Jordan-Young and Karkazis 2019). Modern parenting guides by scientists even still refer to them as "male and female hormone" (Sax 2017).

By 1953, scientists had discovered that testosterone levels influence the genital development of fetuses. Not long after, others argued that the same was true for brains: fetal testosterone levels permanently "organized" brains as male or female, just as they shaped genitals, therefore determining behavior later in life (Jordan-Young and Rumiati 2012; Phoenix et al. 1959). This is the "organizational/activational hypothesis" that underlies most recent research on biological sex differences. Champions of the hypothesis write in terms of "essential difference" and "sexual dimorphism" (literally "two forms"). They discuss how men and women's brains are "hard-wired" differently, creating a sense of clean, categorical distinction that, like the name "sex hormones," is far from the biological reality (Baron-Cohen 2003; Joel et al. 2015; Leslie 2019). Nearly every work in this genre admits that no such clean, categorical distinction exists in human brains or behavior. Authors include illustrations of two overlapping bell curves and admonitions that average differences between men and women should not be read to mean all men or all women are one way or another. Baron-Cohen goes so far as to say that individual women may have "male brains."

As Gina Rippon points out, however, readers "may not hover too long on the semantic niceties of a 'male brain' not meaning 'the brain from a man'" (2019:240–41). Sex difference scientists do not hover long on them either. In *The Essential Difference*, Baron-Cohen introduces brains as a spectrum from those that are good at systematizing (type S) to those good at empathizing (type E), with a plurality of brains falling statistically in the balanced middle (type B) (2003). Within a few pages, however, he switches to calling type S "male brains" and type E "female brains," disregarding his own assertion that the types do not correspond neatly with male and female people. Balanced brains, supposedly the most common type of brains, are entirely absent from most of his discussion. In keeping with the title, readers are left with a sense of *Essential Difference* between men and women. As with hormones themselves, complex brain biology that affects both men and women gets recast in terms of simple, categorical, inborn difference between male and female. Such is the internal logic of the sex difference paradigm.

### 3.2 It Is Known

For the most part, none of this history appears in writing from proponents of sex differences. Science writing generally does not go into the history of its field. Sex difference research is no exception: most papers and books focus on recent advances and current knowledge, leaving history of science as an altogether separate discipline. Writing about sex differences, Unger and Dottolo observe that "history is not highly regarded by psychology because of the field's commitment to… 'just the facts'"(Shields 2016; Unger and Dottolo 2016:278). Consciously or otherwise, authors write centuries of controversy, and their position in it, out of sex science. In so doing, they perform the "god trick" of appearing to have a "view from nowhere" (i.e. with no social history or agenda), which lends their work scientific credibility by making it seem objective (Haraway 1988). Since all research is influenced by the motives, perspectives, and assumptions of researchers, Sandra Harding refers to this as "weak objectivity" and argues that explicit engagement with the social dimensions of scientific work—

reflexivity on the part of researchers—can produce better science (1993). Without reflexivity, flawed science and revisionist histories may flourish.

More insidious than the general lack of historical reflection, however, is the elision of specific, germane controversies. Take, for example, Alexander and Hines' study of vervet monkeys (2002). The authors concluded that monkeys, with no human gender socialization, showed gender-typical toy preferences that mirrored human children. Therefore, they argued, there must be some biological, innate component to differences in interests between human men and women. The study has been critiqued numerous times for including confounding variables; for downplaying its own no-difference findings; for using toy gender labels that are inconsistent with the explanations offered; for presenting the toys in a nonstandard way; and more (Eliot 2009:108; Fine 2010:124–26; Jordan-Young 2010:234–36). Given how gender labels were assigned to toys, the results also contradict the only similar study published to date, which used rhesus monkeys (Hassett, Siebert, and Wallen 2008). The vervet study is ubiquitous in reviews of sex difference research (Bao and Swaab 2011; e.g. Hines 2011; Sax 2017; Soh 2017). Yet those who cite the study rarely mention that it is contested. Even the original authors—who are aware of critiques—present their findings with less ambiguity over time (cf. Hines and Alexander 2008). Of the "feminine" toys, a cooking pot and a doll, the pot is downplayed. The confounding color variable disappears, along with males' equal preference for "masculine" and "feminine" toys. And the rhesus study is cited as corroboration rather than contradiction (Hines 2004, 2011).[16]

Many aspects of this are normal in science. Researchers publish new ideas and results. Replies and critiques routinely follow. Subsequent references to work are generally simpler than initial reports of it. In the field of sex differences, these simplifications led to a false sense of consensus among studies with conflicting methodology (Bleier 1986; Jordan-Young 2010). Sometimes, initial scientific findings turn out to be unsupported by follow-up research. Meta-analyses have shown that this is very common in the sex difference literature (Hyde 2005). It is a core feature of science that we sometimes publish incorrect or contested conclusions. Science is a process, not an infallible dogma.

This becomes problematic, however, when critical engagement is ignored and findings are presented as if they were universally accepted. Consider retractions. Papers are retracted in

[16] To her credit, Hines has revised her position on other branches of sex difference research after years of methodological critique (Leslie 2019).

every discipline for many reasons, ranging from benign statistical errors to gross misconduct and data fabrication. Yet an analysis of 1,775 retractions found that retracting papers does little to stem the flow of citations to those papers, and that the vast majority of citations to retracted papers cite them as if they had not been retracted (2013). Unfortunately then, criticizing, correcting, or even retracting bad research is not enough. Scientists in general, and sex difference scholars in particular, need to engage with the historical context and debate around their sources in all of their work if they are to avoid perpetuating inaccurate information or settle methodological disagreements (Shields 2016).

Unfortunately, some sex difference scholars are actively hostile to historical perspectives. Two critical reviews of feminist neuroscientist Gina Rippon's recent book (2019) are instructive:

> Rippon also builds her case with historical examples of "neurosexism". One shockingly offensive example she quotes is from the anthropologist Gustave Le Bon, who wrote in 1895: "Women… represent the most inferior forms of human evolution…" However, Rippon goes farther still. She argues that… scientists are perpetuating such historical sexism in a new guise (Baron-Cohen 2019).
>
> A book like this is very difficult for someone knowledgeable about the field to review seriously…. Suffice to say it is replete with tactics that are now standard operating procedure for the anti-sex difference writers…. tactics include… resurrecting 19th century arguments almost no modern neuroscientist knows of, or cares about (Cahill 2019).[17]

Both reviews refuse to engage with the substance of historical comparisons offered by feminists.[18] Bringing the history of sex difference research into contemporary scientific discussions is framed as "offensive" and irrelevant. This hostility to discussion of history is what Nancy Tuana calls an "epistemology of ignorance" (2004).

Tellingly, Cahill admits that neuroscientists do not know the history of their own field or the claims they make. Similarly, Baron-Cohen recounts a "recent" revelation:

> Professor Konrad Lorenz [is] widely regarded to be the founding father of ethology, and the master of careful behavior observation and

[17] These reviews were published in right-leaning venues: The Times (London) and the "intellectual dark web" blog Quillette. Positive coverage of Rippon's book appeared both in left-leaning venues like The Guardian and also in Nature.

[18] A a long form of this argument appears in (Gross and Levitt 1994).

measurement…. I read his books at the tender age of nineteen…. A recent [2001] book points out that, despite his high intelligence, the esteemed Lorenz was unable to see that the political ideology of ethnic purification in Germany in the 1940s where he worked, and indeed his own views on eugenics, were hurtful and even dangerous (2003:27).

Lorenz was a Nazi who defended his eugenicist beliefs and research long after the war (Lorenz 1973). This revelation did not prevent Baron-Cohen from writing a glowing apologia for Lorenz. Indeed, Lorenz is brought up only because he is an example of the "male brain," which is brilliant at systematizing but poor at empathizing (and thus prone to being both a great scientist and a Nazi, with no apparent conflict between the two). Some sex difference scholars, then, are not simply unaware of the political and social history behind their ideas; they are also unfazed by learning about it. Most are not so explicit. Hines, for instance, has read and replied to books by Fausto-Sterling and Jordan-Young. But her replies ignore their lengthy, well-documented historical arguments as if they were irrelevant to modern scientific questions (Hines 2011).

### 3.3 The Scientist as Recently Liberated

When sex difference scholars do discuss the history of their field, many tell a story of overcoming oppression. Nowhere is this more apparent than in Alice Dreger's *Galileo's Middle Finger*. Dreger begins by building up her progressive credentials: she calls herself a "liberal feminist," recounts her support for intersex activism, and celebrates the work of "Marxist and feminist science-studies scholars" like Hubbard (2015:4). Quickly, however, readers learn that she has been condemned and pushed out of the academy by the powerful forces of "PC culture," feminists, and trans activists for seeking "dangerous" scientific truths, supposedly just as Galileo Galilei was persecuted for his research. Such 'truths' include defining transgender as a paraphillia and defending a biological drive to rape with evolutionary psychology.[19] In response to critics, Dreger describes her own "reactionary" desire to "make a point of studying … *race and IQ*," which she admits can do "no good and much harm," just "in order to prove how important truth seeking is" (2015:132–33 emphasis original). In the end, she bucks the oppressive forces that would silence her and forges her way as an independent scholar.

Baron-Cohen tells a similar story. In the first pages of *The Essential Difference,* we read that he "would like to believe that, deep down, men and women's minds do not differ in essence"

[19] For critique, see: (Bleier 1997; Serano 2008; Travis 2003).

and he "remain[s] a staunch supporter of efforts to eliminate inequality in society." He was hesitant to write because,

> Discussing sex differences of course drops you straight into the heart of the political correctness debate.… The topic was just too politically sensitive to complete in the 1990s. I postponed finishing this book because I was unsure whether a discussion of psychological sex differences could proceed dispassionately…. My women friends, most of whom consider themselves feminists, have persuaded me that the time is ripe for such discussion (2003:11).[20]

Cahill echoes the same sentiment. He stresses his progressive commitment to including women in medical research, so that treatments do not have unforeseen adverse effects.[21] Yet he received "strong advice to steer clear of studying sex differences from a senior colleague around the year 2000…. For the vast majority of his long and distinguished neuroscience career, exploring sex influences was indeed a terrific way… to become a pariah in the eyes of the neuroscience mainstream" (2017:12).

Dreger, Baron-Cohen, Cahill, and others deploy a common narrative: despite their liberal and feminist beliefs, their scientific pursuit of objective truth has led them to insist that men and women are innately different in their abilities and desires. Those who disagree with them are a "cult" of "extremists" with "deeply ingrained, implicit (but false) assumption[s]" (Baron-Cohen 2019; Cahill 2017:12; Pinker 2002:x). Sex difference scientists, we are told, pay a steep political price for telling their uncomfortable truths.

Interestingly, feminist biologists tell essentially the same narrative, but in reverse. Hubbard's is representative of many early feminist biologists' accounts:

> That I was able to turn my attention to these issues was due to the fact that in 1973, owing in large part to the political work of the women's movement, the tenuous position I had held at Harvard became stable. In an unusual step, the university promoted a few of us from the typical women's ghetto of "research associate and lecturer" to tenured professorships (1990:1–2).

[20] For an argument that such feminists are detached from the goals, motives, and history of feminism, see (Epstein 2007).

[21] Scholars point to the 2013 US FDA decision to recommend different dosages of sleep medication for men and women as proof that sex is key in medical research. Subsequent research has shown that that decision was misguided, however, suggesting the need for caution rather than zealous pursuit of innate sex differences (Greenblatt, Harmatz, and Roth 2019).

From her newly secure position, Hubbard was able to pursue her own research interests and develop a network of colleagues who would go on to publish and edit some of the foundational works in feminist science studies. Changing culture and social movements led to new-found academic freedom, which enabled a career doing otherwise unpopular work. By the mid-1990s, mainstream biologists remarked with horror that "this [feminist, leftist biology] literature grows with astonishing speed" and "that the only widespread, *obvious* discrimination today is against white males" (Gross and Levitt 1994:108, 110 emphasis original).

Still, feminist biology was a tenuous field. Many of its major figures transitioned from science into women's studies, philosophy, or history departments. Many of its publications were in humanities-focused journals like *Signs* and *Hypatia,* or in books and edited volumes. Not all, of course. But it is harder for feminist biologists to make it in science than for sex difference scholars, Fausto-Sterling argues. In response to claims that it just recently became safe to study sex differences in 1992, she pointed out that she wrote a whole book about sex difference research from the 70s. "With few exceptions," she says, "scientists who have taken a different road have a far more difficult time. Their work is less well-known and certainly receives less press coverage, they have a harder time finding jobs, and they often end up working in less prestigious schools, making it harder to get grant money" (1992b:258). Such material disadvantages suggest a real cost to doing critical feminist work in the life sciences, in direct contradiction of sex difference scholars' claim that feminists set and police research agendas in the field.

So which narrative is better supported by the evidence? The list of more than two dozen biologists, endocrinologists, geneticists, molecular biologists, neuroscientists, psychophysiologists, and zoologists in this chapter's opening shows that critics of sex difference are not anti-science. They believe that the biological aspects of sex and gender are worth studying and important for non-scientists as well (Fausto-Sterling 1992a; Hubbard et al. 1993; Jordan-Young 2010:10). But for all their intellectual care and success, feminist biologists have never reigned supreme. Politics before 2000 did not prevent the publication of sex difference research. More than that, "sex differences" has always been a more successful, dominant field of research than feminist biology. In 1997, when Baron-Cohen and Cahill felt unsafe speaking about sex differences, the author of *Men are from Mars, Women are from Venus* boasted that the book had "sold more than ten million copies" and was "a bestseller in more than 40 languages"

(Gray 1997:xii). Indeed, Baron-Cohen credits another pop-science best seller from 1989 for his core idea about "brain sex" (1999).

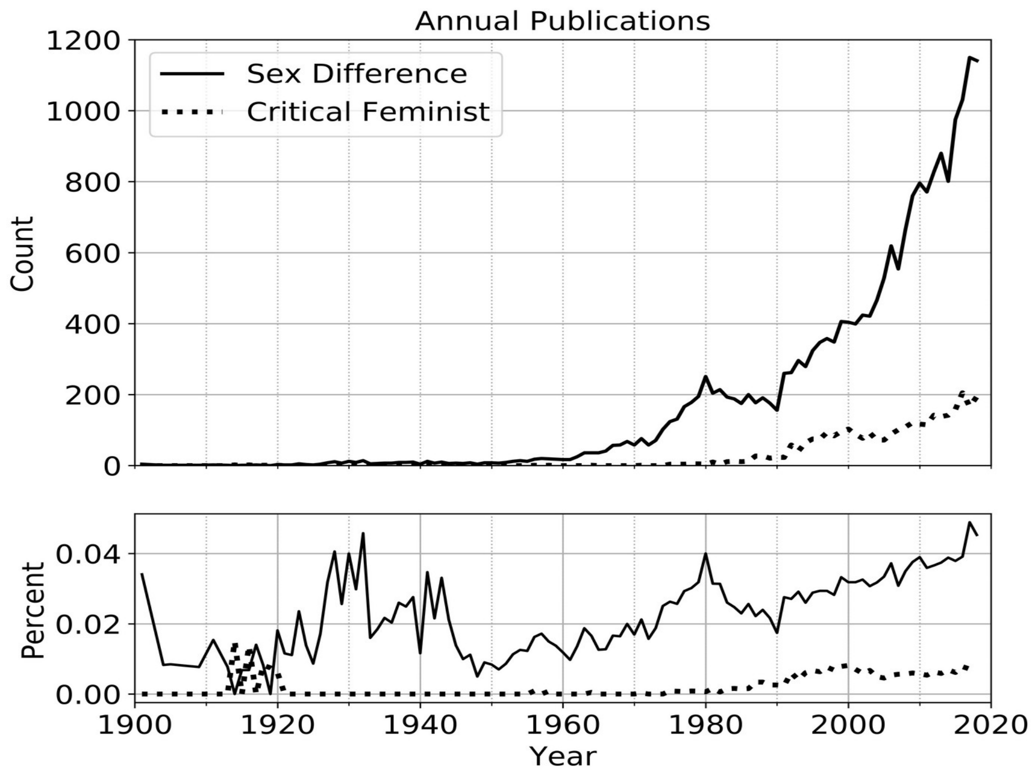


Figure 5: Publications in sex differences and feminist biology, 1900-2018.

Figure 5 shows the volume of research about sex differences and critiques of it published from 1900-2018.[22] It is clear from the publication counts in the top panel that those in favor of sex differences have always dramatically outnumbered critical feminist publications. At no point were critical perspectives dominant.[23] The total number of scientific publications has grown exponentially over time, so the rapid growth of sex difference publications should be read with caution. The lower panel shows the same data as a percent of all publications in a given year. From it, we can see that sex difference research has been a part of English language academic

[22] The top panel shows absolute counts, while the bottom panel shows the same data as a percent of all publications in the Web of Science Core Collection. Sex difference publications are counted as those with some variant of "sex difference," "sex[ual] dimorphism," or "[fe]male brain" in their title or abstract. Critical feminist publications are counted as those with some variant of "[sex/gender] similarity"; "feminis[t/m]" and also "[biology/science]"; or authored by any of a set of feminist critics. Medical and animal-only publications are excluded. If a publication matches both searches, it is counted as critical feminist only. This approach is a conservative estimate of sex difference publications' dominance: including medical and animal publications doubles the gap; and adding difference authors, removing feminist ones, or including "gender difference" widens it as well.

[23] In 1914, 1916, and 1919, there are 1-2 feminist papers, compared with 0-1 difference ones, out of approximately 13,000 papers each year. Over 5-year intervals, difference publications in this period are dominant.

publications throughout the last century. It was particularly common from 1925 to 1945, and has generally grown as a share of all publications since 1950. At its lowest points in the last 100 years, sex difference research made up a similar proportion of all scientific publications as feminist biology did at its highest points.

These data fit well with the histories told by feminist biologists: social change and feminist movements brought more women into the academy and offered some of them tenure in the 70s and 80s. Newly secure in their positions, it became safer and more feasible to publish critiques of sex difference literature. These early publications and career advances paved the way for subsequent generations of feminist biologists. More than that, they correspond with a ten year period of decline in publications about sex differences. This may be where some difference scholars' sense of persecution comes from. Their field really was in decline during the 80s, while feminist scholarship was on the rise. Pro-difference papers held a near monopoly before 1980, when they made up 98% of all publications on the topic. Their market share fell to just 79% of new papers in 1997, when the share of feminist critiques peaked. Still, sex difference scholarship has always been dominant, and scholars like Melissa Hines, who published sex difference research in the 80s, do not, then or now, tell the stories of being recently liberated or persecuted by "PC Police" that later generations of difference scholars tell.

### 3.4 The Big, Bad Social Constructionist

When they are not revising history to erase all critique or to paint themselves as the victims of powerful critics, some sex difference scientists tell a cautionary tale about the dangers of disagreeing with them. This conveniently sidesteps the need to engage with gender socialization research. It is the tragic "John/Joan Story," about a clinical patient whose real name was eventually revealed as David Reimer. Shortly after Reimer was born in 1965, his penis was destroyed in a botched circumcision. John Money and Anke Ehrhardt counseled his parents to raise him as a girl and oversaw his care. They published claiming wild success in socially reassigning gender. As he got older, however, Reimer transitioned back to living as a boy, had his penis surgically reconstructed, and eventually married a woman. He died by suicide in 2004. As sex difference proponents tell it, "the irreversibility of programmed gender identity is clearly illustrated by the sad story of the John–Joan–John case" (Savic, Garcia-Falgueras, and Swaab

2010:44). Indeed, this seems like a clear-cut case of biology trumping socialization: Reimer was born a boy, and despite many efforts to socialize him otherwise, he insisted that he was a boy.

The John/Joan Story is used for more than demonstrating the importance of biology and the impotence of socialization for gender, however. Sex difference proponents use it to argue that those who emphasize gender socialization are not only *wrong*, they are actively *harmful*. As Baron-Cohen says, "John Money, the infamous paediatrician of the 1960s, ignored biology at his peril, in claiming that a child's gender could be determined purely by experience…. Tragically, this dishonest sex reassignment recently led to suicide" (2007:92). The implication is that the people who disagree with them are not simply defying Biological Truth; they are *causing misery and suicide*. Money is figured as the quintessential big, bad social constructionist, the villain in a cautionary tale.[24]

The team that led Reimer's care acted unethically. They reported wild success in changing his gender in the academic literature for years when it was clear that Reimer was deeply unhappy with his gender assignment (Davis 2015; Eliot 2009). Worse, their efforts to socialize Reimer into a girl were highly traumatic, including medical deception, frequent physical and psychological examinations about his sex, and "simulated" sexual intercourse with his twin brother (Colapinto 2000; Jordan-Young 2010). Many aspects of his treatment, including medically unnecessary surgery on children too young to consent, withholding medical information from patients, and frequent sexual examinations, are things intersex activists have campaigned against (Davis 2015). Even though these details could be used to further demonize Money, they are left out of essentialist accounts. Reimer's traumatic and highly unusual childhood means that the John/Joan Story is not representative of research on gender socialization or the social construction of gender. As a response to that work, it is a strawman. In order to lump their critics together and pit "biology" against a "socialization" boogeyman, sex difference proponents tell a selective history. There are also multiple cases of other babies like Reimer who had more positive outcomes, but such cases get less attention in the media and are generally omitted from sex difference proponents' accounts.[25] Perhaps most surprisingly, the accounts demonizing Money for denying biology leave out the fact that Money and Ehrhardt

[24] Ehrhardt was a prolific scholar on these issues with and without Money, but she is often left out due to the Matilda Effect (Rossiter 1993).

[25] Positive outcomes are reviewed in (Meyer-Bahlburg 2005). These outcomes should not be read to undermine the important critiques of unethical practice raised by intersex activists and scholars (Davis 2015). I mention them only to demonstrate the highly selective telling of the John/Joan Story.

went on to become major proponents of brain organization theory and increasingly dismissive of social factors (Bleier 1986:150–52).

Beyond the John/Joan Story, proponents of sex differences have almost no engagement with the vast array research on the social aspects of sex and gender. Sociologists have long studied socialization and the social construction of sex and gender. Indeed, "Sex and Gender" is the largest section of the American Sociological Association, with more than 1,100 members. Yet one is hard pressed to find any references to the relevant sociological research in sex difference publications (Risman 2001). Even in books with sprawling 28- and 44-page reference lists, Baron-Cohen and Hines each cite just a single article from a sociology journal and each cite only eight sources with a predominantly sociological argument (Baron-Cohen 2003; Hines 2004). By comparison, Baron-Cohen cites publications where he is first author 29 times, and Hines does so 19 times. Baron-Cohen's chapter on "Culture" is primarily populated by citations to evolutionary psychologists who are critical of cultural influence.

The situation in most journal articles, where space is more limited, is bleaker still. Many simply ignore social influences on gender. Dick Swaab and colleagues frequently assert that "there is no proof that the social environment after birth has an effect on the development of gender identity" (Savic et al. 2010:41). In support of this claim, they offer a single citation to Simon LeVay's controversial 1990s research. Others point to a 1991 analysis to argue that there are minimal differences in how parents treat boys and girls and then move on (Lytton and Romney 1991). Yet the same 1991 study has also been cited to show the opposite, and to clarify that similar treatment of boys and girls by parents happens only when children conform to gender expectations (Eckes and Trautner 2012). These latter interpretations are supported by the sociological literature (e.g. Field and Mattson 2016; Kane 2006). Moreover, a parent-only focus ignores the broader social milieu. Children are exposed to messages about gender from parents, yes, but also from siblings, peers, teachers, and coaches; from books, movies, and television; from sex-segregated activities, clothing, and toy store isles; and from myriad other sources. As children and adults participate in a gendered world, we do not simply absorb outside influence: we actively participate in constructing group differences for any arbitrary groups we find ourselves in.[26] When scientists fail to engage with the extensive research on how social processes

[26] For a partial review of these complex processes, see (Gansen and Martin 2018).

influence gender differences among people, they present a distorted view of the relevant, empirically grounded research on sex and gender.

### 3.5 Conclusions

I have focused on common examples of historical revisionism within sex difference research, but it is important to note that these patterns are not universal. Like their feminist critics, the proponents of sex difference research are a heterogeneous group with varied agendas and arguments. Some, like James Damore, Anne Moir, John Gray (author of the *Mars and Venus* series), Leonard Sax, and Debra Soh, write for a non-scientific audience. They make explicit personal, social, and policy arguments about hiring fewer women in technology firms, single-sex schools, division of household labor, or parenting style. Most scholars who publish academic work on sex differences, however, ritually distance themselves from such prescriptive, deterministic writing by telling readers, "the view that men are from Mars and women Venus paints the differences between the two sexes as too extreme" (Baron-Cohen 2003:9; Cahill 2019; Hines 2004). Statements like these position them as part of the "reasonable middle" between extreme biological or cultural determinism. They frequently raise the same social and policy questions as their motivation, claiming they want to shed light on those debates. But after many pages arguing for innate gender differences and the natural inevitability of gender inequality in work, aggression, and caring, they stop short of answering social policy questions, leaving readers to infer the rest based on stereotypes.[27]

Sex difference scientists' motivations are varied. Some, like Simon Baron-Cohen, express worry about oppression and denigration of men. He says that "hopefully, in reading this book, men will also experience a resurgence of pride at all the things they can do well," a list which includes "the most wonderful scientists, engineers,… bankers," and "even lawyers," compared to women's "primary school teachers, nurses,… or personal staff" (2003:184–85). Scholars like Baron-Cohen are primarily concerned with differences in *abilities*, what men and women are good at and so what roles in society they should fill. In contrast, scholars like Melissa Hines are explicit that they do not see socially meaningful differences in abilities, but rather in *preferences*. Noting the well-documented placebo effect and stereotype threat influences on performance, she admonishes her colleagues: "reports that hormones cause girls or boys to perform more poorly in certain areas or limit their occupational prospects, even when erroneous,

[27] For a damning critique of this writing style, see (Bleier 1986).

are not benign," because such reports can cause the very outcomes they claim to describe (2004:228). For Hines, differences in play behavior and occupational outcomes result from innate preferences for certain kinds of activity, preferences we share with our monkey relatives, not from women's lack of ability or from social influences.[28] Still others, like Larry Cahill, say they are motivated by a desire to ensure medical treatments are tested on women before they are approved for treatment of women.[29] As always, some scholars confound simple categorization. Alice Dreger and Sara Blaffer Hrdy, for example, have each defended and also critiqued sex difference research on both scientific and social grounds.

Almost no one in or adjacent to the scientific community has argued for the outright superiority of men over women in the last few decades. Sex difference scholars ritually invoke the refrain that "overall intelligence is not better in one sex or the other" in order to show that they, and science, are not sexist (Baron-Cohen 2003:10). Superiority may exist in particular abilities or interests, but overall men and women are "equal," just not "the same" (e.g. Cahill 2017; Pinker 2002). Amusingly, the scientific "truth" that men and women have equal average intelligence is a deliberate choice on the part of intelligence scientists. It is "socially constructed" in the most straightforward way. Intelligence testing and measurement was long controlled by eugenicists for the purpose of demonstrating the superiority/inferiority of ethnic, class, and gender groups (Gould [1981] 1996). Later tests were revised to minimize group difference rather than establish it. Now, questions on intelligence tests are included either if they show no gender difference in performance or if they show a small difference that can be balanced out by another question. Questions showing large gender differences are thrown out (Halpern 2012; Wechsler 1958). Scientists have "the ability to construct valid measures of intelligence that would favor either sex," but deliberately chose to find sameness instead (Hines 2004:211). So, if sex difference scholars are right, and modern science is "not sexist" because it finds that women and men have the same general intelligence, then *modern science is not sexist because scientists chose not to look for sex differences.*

Opposition to sex essentialism in scientific research is not made up of powerful, anti-science ideologues out to enforce "political correctness." Rather, feminist biology is made up of those who care deeply about and thoughtfully engage with research on sex and gender. They call

[28] For critique, see (Cech 2021).

[29] This motive does not explain his vitriolic review of Rippon's book, which is about the social rather than medical implications of biological sex. For more on sex difference motives and meanings in medicine, see (Epstein 2007).

on their colleagues and the public to avoid the scientifically unsound rhetoric of essential, innate, and categorical differences and the socially harmful effects that rhetoric has. The solution is more and deeper engagement with the science of sex and gender, not less. And that engagement must include a fuller, more accurate picture of the field's history and citational practices.

## Chapter 4 Experiencing the Work: Paradigms of Sex Research and Women in STEM[30]

In 2005 Harvard University's president, Lawrence Summers, sparked heated debates over the causes of women's under-representation in many STEM fields. He claimed scientific research led him to believe that "intrinsic aptitude" was more important than "socialization and continuing discrimination" (2005). That is, women are less likely to be biologically capable of STEM work than men. While Summers' remarks were a flash point of widespread public attention, the views that they reflect are a longstanding fixture in scientific research on sex. The tradition of sex difference research stretches much further back than Summers' suggested date of 1990, and it continues to guide publications and inflame scientific debates today, with over 1,100 academic publications on sex difference in 2017 alone (Lockhart 2020; Sanz 2017). Scientific claims of women's inferiority are not new. Nor, sadly, are they old.

It is axiomatic among sociologists of science that the content of scientific work is not independent from the social lives of scientists. Typically arguments focus on how scientists' social positions influence their work. This chapter explores evidence for influences in the opposite direction. The content of scientific research may influence who becomes a scientist, because the content of scientific work is a part of the professional culture of scientific fields. Characteristics of "good science" like dispassionate objectivity get transposed into characteristics of "good scientists," often in ways that discredit women and other people with marginalized identities (Cech 2013b; Haraway 1988; Subramaniam 2000). The content of science can both reflect and influence the "schemas of inequality" within a field, which in turn affect selection, persistence, evaluation, promotion, resources, equity initiatives, and more (Cech 2013a). This process may be especially clear in research on the biological dimensions of sex, where scientists openly debate the abilities, preferences, and ultimate merits of men and women in both the content of their research and their professional career interactions.

To be sure, there is an extensive literature on women's (under)representation in science and in occupations more generally. A google scholar search for "women in science" returns 4.5

[30] A version of this chapter has been previously published as Lockhart, Jeffrey W. 2021. "Paradigms of Sex Research and Women in STEM." *Gender & Society* 35(3): 449-475.

million results. As Nancy Brickhouse notes, this research has "provided us with ideas about what could be done to change the domination of men in science and science education. However, this research did not address epistemology and thus left unquestioned traditional conceptions of what counts as knowledge" (2001:282–83). Next to findings about culture, discrimination, educational tracking, ideal worker norms, and more, the contents of published sex research are not the only or the largest influences on women's participation in science. Nevertheless I use large-scale quantitative analysis to argue, along with other feminist scientists and feminist science studies scholars, that the way scientists write about sex matters (Hubbard 1990; Martin 1991; Sanz 2017).

Specifically, I propose and test a link between the content of scientific research on sex and women's subsequent representation in 53 subfields of the life sciences over the course of 47 years. I draw on newly available historical data from the National Science Foundation's Survey of Earned Doctorates (SED) on the prevalence of women in detailed academic subfields from 1970 to 2017. I match this data with the Web of Science Core Collection (WoS) database, containing 69 million academic publications. I show that the amount of publications positing innate, categorical differences between men and women in biological and health science subfields is negatively related to the proportion of women earning PhDs in the same subfield in future years. To my surprise, however, I find the opposite effect in psychology subfields. Furthermore, I show that alternate paradigms of sex research in the same life science fields have different effects. Critical feminist research that challenges biological essentialism is associated with future increases in women's graduation rates, but research narrowly aimed at including women to improve their health outcomes has no effect. Perhaps more STEM research should take a critical feminist approach.

### 4.1 Paradigms of Sex Research in the Life Sciences

In order to understand the effects of how life scientists write about sex, it helps to know what they say. Today, there are three main paradigms for making sex the object of study: S*ex Differences*, *Feminist Biology*, and *Gender Medicine.*

**Sex differences** research approaches sex categorically, as male or female, and seeks biological explanations for observed differences between them. Much of this research has focused on cognitive abilities and behaviors, leading to questions like whether "female brains" are capable of or interested in STEM employment. Although the biological specifics of sex difference research have changed considerably over time, this core approach—essentializing differences between categories—has been relatively stable (Lockhart 2020; Sanz 2017). Critics label some of this work "neurosexism" (Bluhm 2021; Fine 2013). If research is sexist, it may disproportionately attract and retain men. Not all women scientists see this paradigm as sexist, however. A number of women have made careers on conducting sex difference research.

This research has a large cultural presence outside academia. It showed up in news coverage of Summers' 2005 remarks and in James Damore's 2017 "Google memo" against hiring women engineers. Steven Pinker and Charles Murray are vocal proponents of the paradigm in popular media and venues like the American Enterprise Institute. Sex difference research is a perennial subject of boundary work in professional sports, determining who can compete in women's events and leagues (Davis 2015; Karkazis and Jordan-Young 2015; Pielke and Pape 2019). It has become especially popular recently with "trans exclusionary radical feminists," who cite biological sex differences as justifications to deny trans people's demands for civil and medical rights. Such scientists form organizations, give invited talks, and serve as expert witnesses (e.g. Hilton and Whyte n.d.; World Rugby 2020).

Social scientists have critiqued the sex difference paradigm repeatedly for misunderstanding the social dimensions of gender differences (Quadagno 1979; Richardson et al. 2020; Tuana 1983). The causal direction of biological-social correlations is not obvious: observed physiological and cognitive differences are sometimes caused by social factors, such as the stress of being a token in the workplace or practiced repetition of gendered activities (Fausto-Sterling 2005; Taylor 2016). In 2000, when J. Richard Udry advocated the sex difference paradigm in *American Sociological Review*, Barbara Risman's reply noted that his article was published "without citing or directly engaging the concepts, arguments, and findings of the considerable literature on the sociology of gender that has been developed over the past 30 years" (2001:606). This is true of the sex difference literature in general: it makes social claims without engaging the social science literature (Lockhart 2020).

There has long been feminist resistance to the sex difference paradigm (Herschberger 1948; Lowie and Hollingworth 1916; e.g. Poulain de La Barre 1677). In the 1970s, buoyed by their local feminist communities, newfound academic job security, and the success of a burgeoning community of feminist primatologists, a group of women in biology began to coalesce around a new paradigm of sex research: **feminist biology** (Hubbard 1990). These scholars sought to build “two way streets” between feminism and biology, arguing that biological research on sex is improved scientifically and culturally by replacing essentialist dichotomies with increased complexity (Fausto-Sterling 1992a; Hubbard et al. 1993). They typically argue that differences between two categories are a poor description of the underlying biology. Many aspects of sex (such as hormone levels) are better measured as continuous variables with complex interactions and social/environmental influences. Concepts like “female brains” are misguided holdovers from a sexist past. They also tend to critique it on ethical grounds, arguing that seeking to prove superiority or justify patterns of inequality is not a good use of our energy compared with alleviating human suffering and expanding opportunities for everyone (for this argument in sociology, see Meadow 2013). Feminist biologists made big contributions to feminist standpoint theory and to critiques of rationality (for review, see Sanz 2017). They have institutionalized over time, with edited volumes, readers, mailing lists, conferences, and labs.

In the 1990s, a third paradigm of sex research rose under the banner of “gender-specific medicine” and “gender biology” (Epstein 2007; Roy 2012).[31] This work calls for greater inclusion of women as subjects in medical research, arguing that treatment and diagnosis of women risked unforeseen harms if medical research remained focused on men. Some of the early research findings that popularized this paradigm, such as prescribing different dosage of sleep medication for men and women, have not stood up well over time (Richardson et al. 2015). Nevertheless, it has been institutionalized with great success. There is a journal and professional society for “**gender medicine**,” and the NIH now requires the inclusion of “sex as a biological variable” in all their funded research. Gender medicine reflects aspects of both sex difference and feminist biology paradigms. It proceeds from an assumption of essential, categorical sex differences, but it explicitly aims to help women and to correct problematic assumptions about sex in research.

Debates among these paradigms are ongoing, with scientists writing both scientific articles and popular media advancing their chosen paradigm. No Kuhnian revolution has taken place. Instead, these paradigms behave more like Panofsky’s “misbehaving science,” reiterating

[31] In the other chapters of this dissertation, I combine the sex difference and gender medicine paradigms into one analytic category. In terms of how they conduct and present research, there is very little to distinguish them. In this chapter, I break them out as separate categories because their stated motivations differ in ways that may change their relationship to women in STEM.

old arguments in insular communities without resolving conflicts (Lockhart 2020; Panofsky 2014; Sanz 2017). But the conversation is not entirely stagnant. The biological details that serve as battlegrounds for these debates have shifted over time among subfields like endocrinology, neuroscience, and genetics. This provides the variation for my analysis.

### 4.2 Professional Cultures and the Content of Science

Professional culture is the semi-autonomous system of meanings, symbols, and habits associated with a profession (Abbott 1988; Cech 2013a). Research on professional cultures has demonstrated that STEM fields draw on varying "schemas of inequality" and "merit" that shape the experiences and career trajectories of scientists (Cech 2013a; Cech, Blair-Loy, and Rogers 2018). These schemas are inculcated during graduate education, and they disproportionately impact the success of students in raced, classed, and gendered ways (Cech 2014; Costello 2005). Gendered schemas about "natural differences" between men and women affect who is valued in the workplace (Schilt 2010), and increasing the number of women in an occupation does not necessarily change them (Hochschild and Machung 1989). The schemas by which scientists judge one anther's abilities, preferences, and merit (Lamont 2009), might reasonably be expected to manifest in and be informed by those same scientists' publications about human abilities, preferences, and merits. Typically, professional cultures and schemas of inequality are studied in rich detail with interviews and ethnography, or across broader segments of the STEM workforce with cross sectional surveys. By using the content of publications and an annual census of doctoral recipients, I provide a population level view of the same processes over half a century. There is much research on local processes by which research publications may influence professional cultures and thus overall trends in the gender balance of scientific fields.

From their beginnings in the late 1960s and early 1970s, women's studies, women's health, women's biology—as well as African American studies—were expressly intended, among other things, to attract and retain more women and Black people in academia by changing the content of curricula and research (Johnson 2020; Messer-Davidow 2002). Others have made the same arguments in physics and chemistry: rethinking curriculum and research through a feminist lens not only improves their content, but should attract and retain more diverse science practitioners (Barad 1995; Barton 1997). Simultaneously, proponents of sex difference research have openly resisted calls to bring more women into STEM on the grounds that disparities reflect natural differences rather than social problems (Hubbard 1990). These paradigm fights have always been partly about influencing the demographics of scientists, not only the "objective truth" of biology and sex.

The content of research can exert push and pull effects on individuals. For instance men —even those few who select into women's studies courses—tend to distance ourselves from things labeled "feminism" (McCabe 2005; Pleasants 2011). "Feminist science," then, may both attract women and repel men. At the same time, science that questions women's inherent ability or pushes essentialist views of sex may create a hostile environment that pushes women, intersex, nonbinary, and trans students out (Philosopher 2019). Working in a hostile discipline can take its toll. As Tey Meadow powerfully put it,

> the very publication of studies like the NFSS exacts a particular form of psychic violence – on families, on families like mine, on scholars like me. As a sociologist and a gay parent, I can withstand the accusations of instability and unfitness it levies.... I can sit in rooms at professional conferences and watch my peers and colleagues dispassionately debate the legitimacy of my relationship and the fundamental right of my family to exist…. The question is whether I should have to (2013).

Such research contributes to a field's professional culture, conveying two messages to students and faculty mentors. First, one's ability or worth is a function of their identity. Second, one's comparative merit and innate character are acceptable—even rewarded—lines of inquiry in their field. Such publications advance a particular schema of inequality by naturalizing group inferiority.

Likewise, devaluing critical feminist research can have detrimental effects on students. Carla Pfeffer (2018) describes the experience of being denied a dissertation grant because reviewers did not believe she could "build a career" or get publications in "the most prestigious" journals by studying the cis women partners of trans men. Negative "feedback may carry a particularly shaming valence for queer scholars if they are told, essentially, that projects reflecting their lives carry little broader import or value," and this can make them doubt their "future career prospects altogether" (Pfeffer 2018:310). Research on sex differences does not debate women's right to exist, and in that sense is not the same kind of psychic violence often directed at LGB and especially trans people, and also at people of color, people with disabilities, and other groups' whose merit is routinely debated by scientists. But sex research does nevertheless debate women's fitness and place in the world. In these cases, developing a professional identity consonant with one's field may require extensive and potentially ruinous emotional labor and loss of identity (Subramaniam 2000). Professional identities dissonant with professional cultures lead to decreased success in graduate school for women and underrepresented groups (Costello 2005).

Moreover, people are prone to believing sex research claims. Summers' comments demonstrate that reading essentialist research about sex can lead people to believe that women are less capable of STEM work. Importantly, bio-essentialist research does not need to speak directly to men and women's relative STEM ability to influence beliefs about it. In the question period after his 2005 remarks, Summers' was asked for evidence to support his claims about women's inability to do STEM work. He replied,

> the field of behavioral genetics had a revolution... the discovery that a large number of things that people thought were due to socialization weren't, and were in fact due to more intrinsic human nature, and that set of discoveries, it seemed to me, ought to influence the way one thought about other areas (2005).

Summers had no relevant research to cite. Instead, he pivoted to generic research in a move that Donovan et al. (2019) call "neurogenic essentialism." They find scientific research can have this effect on audiences "by activating and strengthening any or all of the following four beliefs composing neurogenetic essentialism: The *uniformity* of categories [like man and woman], the *discreteness* of categories, the *underlying essence* of categories, and the *causal influence* of this essence on category members" (2019:722). An array of experiments has shown that reading essentialist explanations of sex differences in newspapers or science textbooks increases prejudiced beliefs about women as a whole, as well as stereotype threat and fear of discrimination among women and girls. These effects hold even when the research presented is not about the stereotype measured (e.g. women and STEM) (Brescoll and LaFrance 2004; Ching and Xu 2018; Coleman and Hong 2008; Donovan et al. 2019).

With regard to professional culture, prejudiced beliefs about gender may influence how men and women are evaluated and the overall climate women experience in science or graduate school. Research on the field-specific ability beliefs (FAB) hypothesis finds that the proportion of women in academic fields corresponds with how much practitioners in those fields say success "requires a special aptitude that just can't be taught" (Leslie et al. 2015). The mechanism for this association remains elusive (Bailey et al. 2019), perhaps in part because this work assumes there is a uniformly negative cultural association between "woman" and "genius." Such a relationship has been demonstrated in many relevant cases (Musto 2019; Ridgeway 2011; Storage et al. 2016). However, my analysis of publications shows that beliefs about the relationship between gender and innate ability are not uniform across academic fields or over time. This variation in gender stereotypes predicts women's representation in the life sciences, making it a missing piece in FAB research.

Essentialist gender beliefs may also interact with ideal worker norms that set long hours and high job dedication as expectations for STEM work, contrasting those qualities with bioessentialist notions of women's family preferences. Ideal worker norms have tangible implications for women in STEM, for example by disproportionately tracking them out of research when they have children (Cech and Blair-Loy 2019). Such research may also make scientific ideal worker norms harder for women to achieve in other ways. In fields where a social group's "nature" is a valid object of study, members of that group may be viewed pejoratively as "mesearchers" incapable of objectivity (Haraway 1988; Pfeffer 2018). Or due to the stress and insult of such work, they may have difficulty maintaining the detached collegiality necessary for career advancement (Cech 2013a; Subramaniam 2000).

These findings also point to structural concerns. Sex difference research can influence university presidents like Summers to believe that they should not allocate resources to gender equity initiatives. Essentialist beliefs about sex enable practitioners to continue believing that science is a functional meritocracy, even when presented with substantial gender disparities; these beliefs are a major barrier to gender equality in STEM (Cech 2017; Hubbard 1990; Seron et al. 2018). By affecting fields, this research affects a student's experience whether or not they personally read papers about sex. Further, students match with mentors and departments on the topics they study. This process implicates students' social identities: LGBTQ undergraduates use the topics fields study to understand whether those fields are "queer-free" or "queer-friendly," and thus whether they belong (Forbes 2020). PhD funding in the life sciences is dependent on how valued a student's topic is (Belavy, Owen, and Livingston 2020). Mentor prestige and financial resources are key factors in life science PhD success and retention, outweighing prior academic achievement (Belavy et al. 2020). Thus the success and prestige of sex research paradigms (signaled to academics through publications) may influence graduate students' choice of subfield and persistence by shaping the resources available to them and their mentors. Indeed, some gaps in funding, training environment, and career outcomes between men and women STEM PhD students can be accounted for by variation dissertation topics (Buffington et al. 2016).

Of course, the relationship between publication content and professional cultures is not one directional. As Prescod-Weinstein explains, even if the paradigm she names "white empiricism" is rare in physics,

> What matters is that their arguments are given room to breathe in professional spaces, whether it is publications, conferences, or books.... While many in the community may disagree..., their epistemic agency is recognized as legitimate. Black women speaking up about their experiences with discrimination are simply not offered the same platforms or axiomatic acceptance of their agency in discourses about race and gender/sex" (2020:430).

The culture of a field affects what is published and who is listened to. Aaron Panofsky explains that behavior genetics persists without addressing criticism, because practitioners exist in fields that value their work, while their critics are only influential in other fields to which they are unaccountable (2014). Thus field-specific professional cultures influence the content of published research, and within-field publications are the primary research influence on scientists.

Taken together, the various mechanisms outlined in this section suggest that the kinds of sex research being published shapes the professional cultures of scientific fields in ways that likely influence the gender ratio of PhD recipients. Sex difference research may decrease the relative numbers of women entering a field, while explicitly feminist approaches may increase it.

## 4.3 Data

### *4.3.1 Sources*

The **NSF Survey of Earned Doctorates** (SED) is an annual census of research doctorate recipients in the United States (in 1970, N=28,861. In 2017, N=54,664). It collects self-report surveys and supplements missing information with school records. SED staff created a new public tabulation of the microdata for me which includes counts of doctorates granted by year (1970-2017), detailed subfield of study, and sex ([male, female]). Field classifications are standardized across years. Missing data is less than 1% in all years, and 0 in most years. Small cells are suppressed: in field-years where the number of men or women finishing a PhD is between 1 and 4, inclusive, the exact number and sex ratio are hidden. Example data is shown in Figure 6.

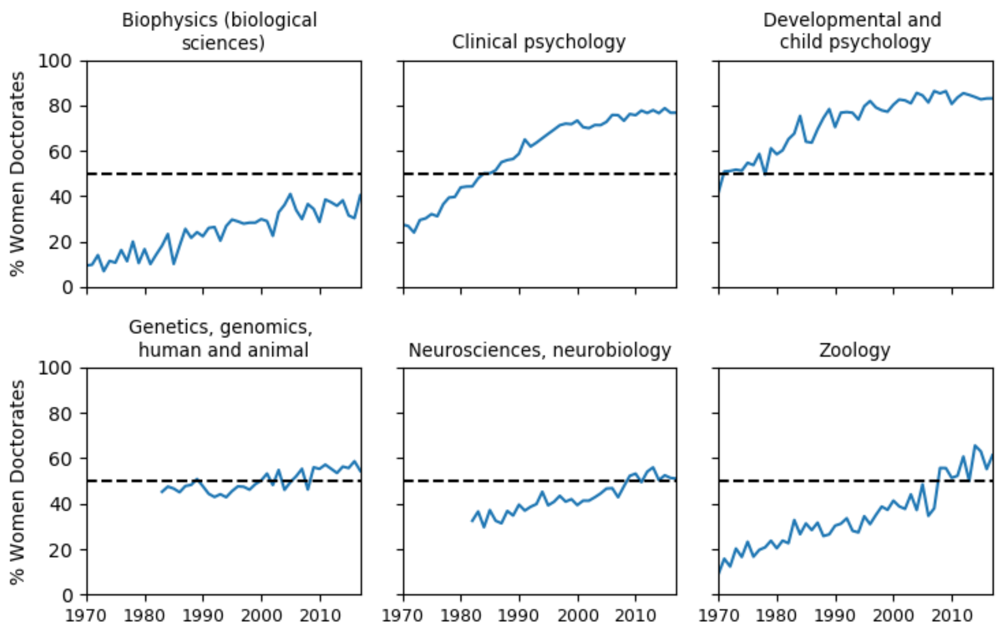


Figure 6: Percent women earning doctorates in select subfields over time.

I restrict my sample to the SED categories "Biological and Biomedical Sciences," "Health Sciences," "Psychology," and "Anthropology," as these are the fields where paradigm battles between sex differences, feminist biology, and gender medicine have primarily taken place. I exclude administrative subfields such as "Health Systems Administration" and those with fewer than 10 years of data. I combine general, cultural, and physical anthropology, which were not measured separately until 2014. I do the same for psychometrics and quantitative psychology. The resulting subfields include some core battlegrounds of sex research (e.g. endocrinology, biological psychology, developmental biology) as well as more removed areas (e.g. microbiology, plant science). They represent the universe of available PhD fields for students interested in the life sciences. The SED data has gender ratio for 1,956 field-years, out of a possible 2,491 (53 fields observed on average 36.9 of 47 possible years).

While most of these fields see increases in women's representation, the rate and timing of growth varies dramatically. Some fields, such as genetics and nursing, see very little change over time. Some start near gender parity, while others start with dramatic inequality. Some subfields achieve or exceed parity, while others never attain it. This paper takes a comparative approach. It is not set up to explain changes in the overall number of women in science or the historical circumstances that have contributed to that. Instead, I compare life science subfields in order to see whether the content of research in those subfields can explain some of why the trajectories of women's participation look so different from one subfield to the next.

The **Web of Science Core Collection** (WoS) is a database of academic publications, metadata, and citations. It has 69 million documents cataloged between 1900 and 2018. I use a copy of the entire raw WoS database, rather than search results from the online WoS interface. This presents a more comprehensive view of research subfields than sampling a handful of top journals, but the WoS database is still biased toward more established and higher ranked journals over books and newer or more obscure journals. I further restrict it to English publications. A substantial amount of feminist science has been published in the journals *Hypatia*, which is not indexed by WoS, and *Signs*, which is indexed under the humanities and not under science subjects. These omissions would limit inferences about feminist science as a whole but not my analysis of within-field effects. Prior work has shown that scientists are held accountable to within-field publications rather than cross-field critiques (Panofsky 2014). Because the database was shared as 735 GB of unwieldy XML files, I used python and spark on one of my university's high performance computing clusters to conduct analysis.

#### *4.3.2 Linking*

The SED and WoS use different taxonomies for research fields. I manually matched the SED fields to their WoS counterparts, consulting the documentation for both. Many were simple (e.g. "Clinical psychology" in SED corresponds to "Psychology, clinical" in WoS). Several subfields of psychology in the SED (namely: cognitive, community, comparative, counseling, family, organizational, personality, and social) were not distinguished in the WoS subject classifications. These I matched on keywords in book and journal titles, manually checking matches to ensure accuracy. In one case I was unable to obtain good matches, so instead I combined the SED categories for plant genetics, physiology, and pathology so that they correspond to the WoS category "plant science." The 53 life-science subfields of the SED matched with 38,541,924 WoS publications.

### *4.3.3 Paradigm Labels*

I designed a set of keyword matches and exclusions to label each publication in the matched WoS sample for which paradigms it invokes, if any. To design and test my labeling, I relied heavily on a parallel qualitative project where I study the rhetorical strategies deployed in sex research paradigm debates (Lockhart 2020). For this analysis, any publication with "feminism", "feminist", "sex similarit", or "gender similarit" in the title, abstract, or keywords was labeled as engaging the feminist biology paradigm, along with anything published in journals or books with "feminism" or "feminist" in the title. "Feminist" and "feminism" had to be specified separately to exclude references to "feminizing." The other terms were included to capture references to Janet Hyde's influential "Gender Similarities Hypothesis" (2005). Explicit references to feminism or gender similarities in scientific writing are hallmarks of the paradigm and they send strong signals of membership to readers. This resulted in 7,195 life science publications (0.04%) being labeled "feminist biology."

Publications were labeled as engaging the sex difference paradigm if they used any of the following in their title, abstract, or keywords: "sex differen", "sexual dimorphism", "sexually dimorphic", or "[fe]male brain", or if they were in a book or journal with a title containing "sex differen". These terms are prolific in research under the sex difference paradigm. I did not include matches on "gender differences", because qualitatively reviewing the matches to that term indicated that most of them were focused entirely on social phenomena without reference to biological underpinnings, consistent with the use of "sex" and "gender" to distinguish between "nature" and "culture." This resulted in 42,901 life science publications (0.25%) being labeled "sex differences."

Finally, publications were labeled as engaging the gender medicine paradigm if their title, abstract, or keywords matched any of these terms: "sex based", "gender based", "gender specific", "sex specific", or "gender medicine", or if they were published in a journal or book with a title containing "gender medicine", "gender specific", "sex specific", or "journal of womens health" but not "international". This search captures the core literature of this paradigm, as described by Epstein (2007). This resulted in 27,646 life science publications (0.16%) labeled "gender medicine." While many of these terms seem general on face, in the context of life science publications, they correspond very closely with the gender medicine paradigm. For example, "sex based differences" is almost exclusively used in the context of clinical health and medicine research, while "sex differences" is the term of choice elsewhere.

Over 12 months I checked and refined the labeling process to ensure that it included important works and excluded false positives such as "feminist science fiction," which matched initial searches for "feminist science." This period also included iteratively standardizing punctuation, spacing, and capitalization in the WoS database to improve text matching.

Importantly, I measure the paradigms represented, not the position of authors or substance of findings. Thus a paper that declares "no sex difference" in some trait is still labeled as a sex difference paper, because it uses the rhetoric of that paradigm and frames research on sex as a question of categorical difference. Similarly, a "feminist critique of sex difference" would be categorized as both feminist and sex difference, since it rhetorically engages both paradigms, regardless of where the author ultimately sides. Overlaps between categories were extremely rare: only 0.17% of feminist biology and sex difference articles overlapped. Overlap between feminist biology and gender medicine was 0.27%, and overlap between sex differences and gender medicine was 6.0%. This suggests that my classification schema achieves a good separation of the paradigms, despite potential ambiguity in the rhetorical landscape.

Categorizing work this way necessarily entails simplifying large and complicated literatures. However, these three paradigms are well documented in sex research, and after iteratively refining the labeling process, reviewing random samples of the results, and finding very little overlap between categories, I am confident that my approach aligns well with these paradigms of sex research.

While articles using any of these three paradigms are rare in the life sciences overall, their relative prevalence varies substantially across time and subfields. Figure 7 shows paradigm prevalence in four subfields. In microbiology, there has been little debate over sex research, which makes sense because the subjects of microbiology research include bacteria, fungi, and viruses. By contrast, all three paradigms have been more popular in general psychology, where feminist biology and sex difference publications were equally prevalent throughout the 1990s, but sex difference research was dominant in other periods. Developmental psychology was one

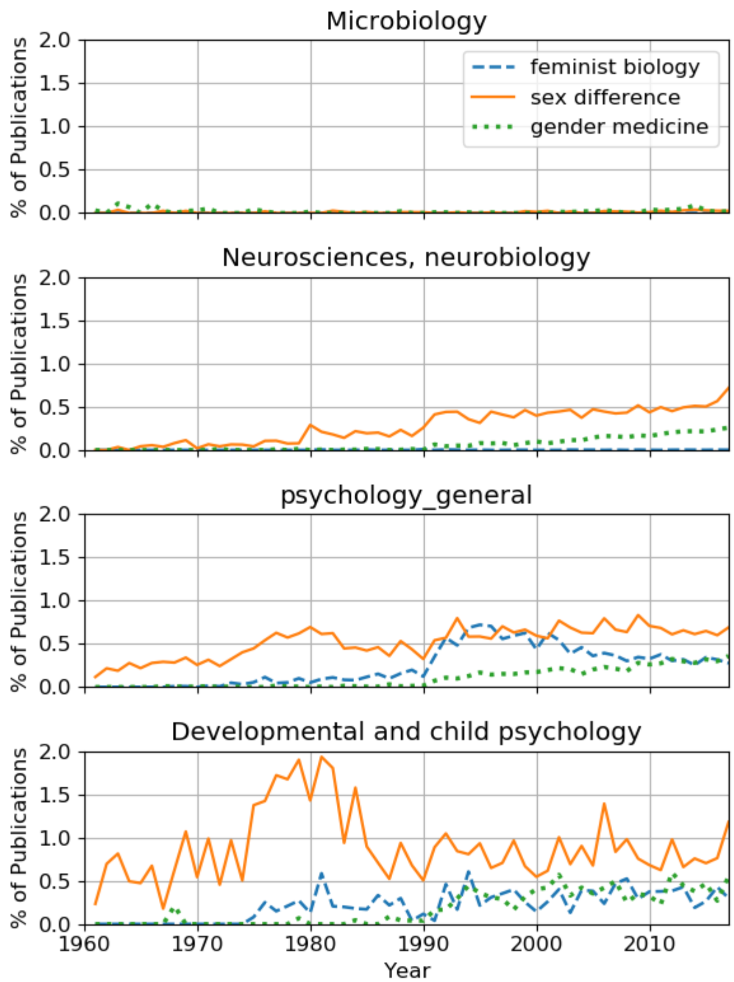


Figure 7: Prevalence of paradigms over time in select subfields

of the early sites of debates around childhood gender, and as expected, it has a greater and earlier engagement with these paradigms.

### *4.3.4 Modeling*

I use linear mixed effects growth models to study the relationship between paradigm prevalence and women's share of earned doctorates in life science subfields. By incorporating subfield-level random slopes and random intercepts, these models allow me to examine variation both within and between subfields in both the overall level and rate of increase in women's participation. Put differently, these models allow the regression lines (trajectories of women's participation) for each of the subfields in Figure 6 to vary, while still leveraging information from the full sample (partial pooling).

**Dependent variable**: Percent of doctoral recipients who are women, by subfield and year, from the SED.[32]

**Independent variables**: for each of the three paradigms, I create a lagged prevalence variable. These variables correspond to the percentage of publications in a subfield, during the previous five years, that were labeled with a given paradigm. For example the variable *feminist biology* for microbiology in 2006 is the percent of microbiology papers published from 2001-2005, inclusive, that were labeled "feminist biology." This lag corresponds roughly to the time period when doctorate recipients would have been graduate students, because undergraduate ability beliefs have been shown not to track with PhD outcomes (Bailey et al. 2019). Since 1970, mean time-to-degree for life science doctorates in the SED has fluctuated between 6.2 and 7.7 years (Laurence 2014).

**Models**: I construct three mixed effects growth models. The first follows Barr et al.'s (2013) advice to "keep it maximal" and include fixed and random effects for all terms. Thus **Model 1** includes both fixed and random effects for subfield, time (years since 1970), and the lagged prevalence of feminist, sex difference, and gender medicine research, along with covariances for all combinations thereof. However, maximal models are often overparameterized: they may fit more terms than the data can support, fail to converge, and be difficult to interpret. Thus I also fit **Model 2**, with fixed and random effects for subfield and time, but only fixed effects for the three paradigms (Bates et al. 2018).

I repeat these analysis using a two-way fixed effects model, **Model 3**, with fixed effects for both subfield and year. Such models are much more conservative, as they remove time trends and between-field variation, leaving only within-field, time-independent effects. That is, they control away much of the variation I am interested in, because I expect that between-field differences in sex research paradigms will influence fields' trajectories over time. If this model shows effects for paradigm prevalence, it will be evidence of a local influence: the prevalence of sex research paradigms in a particular field and time influences the rate of women's graduation above and beyond broader field and time trends.

I use cross validation to test the robustness of my findings to sampling decisions. To do this, I create subsets of the data with each of the subfields omitted. I create additional subsets, with each of the broad fields omitted and with each broad field in isolation. I fit the same model

[32] Unfortunately, NSF surveys of enrolled graduate students use a different set of subfield categories than the SED, so it is not possible to disaggregate the effects of selection into doctoral programs from attrition on degree outcomes.

specifications to all subsets of the data and evaluate whether the inclusion/exclusion of specific fields or subfields substantively changes the results. The cross-validation results suggest an interaction between the broad field of psychology and the prevalence of sex difference research, so I fit **Model 4** with the same specification as Model 2, but with the interaction added as a fixed effect. (In Model 1, the random effects already capture this between-field variation in the effect of sex difference publications.)

Because not all publications are equally influential, I repeat all of these analyses weighting publications by their citations. A publication's weight is set to the number of papers citing it within five years of publication, plus one (the publication itself), logged due to the long tail. These models produce qualitatively similar results but are not shown because "log citation units" are harder to interpret than simple fractions of published articles.

### 4.4 Results

Selected results for Models 1-4 are shown in Table 1. To my surprise, they show stronger and more robust effects for the publication of feminist biology research than for sex difference research. Model 1, the "maximal model" with fixed and random effects for *subfield, time, feminist biology, sex difference*, and *gender medicine*, as well as covariances among them, shows a significant, positive effect of publications that use the feminist biology paradigm on the future proportion of women among a subfield's doctoral recipients. The main effect is large: a one percentage point increase in the amount of feminist biology published during a five-year period corresponds to a 30 percentage point increase in the amount of women graduates! This effect should be interpreted in context: feminist biology comprises under 1% of publications in all but eight field-years. A one standard deviation increase in the prevalence of feminist biology corresponds to a more moderate 4.7 point increase in women's share of PhDs earned.

| N Obs. = 1,956 | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N Subfields = 53 | Coef. | Std.Err. | | Coef. | Std.Err. | | Coef. | Std.Err. | | Coef. | Std.Err. | |
| **Feminist Biology** | 30.22 | 10.53 | ** | 7.77 | 1.54 | *** | 3.18 | 1.41 | * | 7.34 | 1.50 | *** |
| **Sex Difference** | 0.88 | 2.49 | | 1.53 | 0.99 | | -0.01 | 0.70 | | -4.79 | 2.04 | * |
| **Gender Medicine** | -10.07 | 7.83 | | -1.14 | 1.58 | | 2.22 | 0.97 | * | -0.34 | 1.46 | |
| **FB variance** | 581.16 | 63.20 | - | - | - | - | - | - | - | - | - | - |
| **SD variance** | 48.67 | 6.18 | - | - | - | - | - | - | - | - | - | - |
| **GM variance** | 1776.97 | 137.46 | - | - | - | - | - | - | - | - | - | - |
| **Psychology** | - | - | - | - | - | - | - | - | - | 9.58 | 3.88 | * |
| **Psychology x SD** | - | - | - | - | - | - | - | - | - | 7.44 | 2.29 | *** |

Table 1: The effect of research paradigms on the percent of women earning PhDs in the life sciences.

The main effects for sex difference and gender medicine are not significantly different from zero in the maximal model. The random effects for all three paradigms are large, indicating considerable variance in their effects across subfields. As Bates et al. explain, "if the effect of A ... differs reliably between subjects [or fields], uncertainty about A may be so substantial that the main effect of A is no longer significant in a model allowing for random slopes for A" (2018). Model 2 fits a more parsimonious subset of the parameters: it still allows for random slopes and intercepts by subfield and time, but it does not include random effects or covariances for the three paradigms. As such, it assumes consistent effects for each paradigm across subfields. Under that assumption, there is still a significant positive relationship between the publication of openly feminist research and the future proportion of women graduates. The effect appears smaller in this model (7.77 points more women per 1 point additional feminist publications, or 1.21 per standard deviation), indicating that outlier fields which are allowed to vary in Model 1 are pulling the average effect down in Model 2 where the effect is constrained to be the same across fields. A likelihood ratio test and theory both indicate that Model 1 is a better fit of the data (Barr et al. 2013).

Model 3 uses two-way fixed effects (i.e. dummy variables for both subfield and year) to further isolate the effect of the paradigms on women's graduation rates net of trends in time and between-field differences, providing a conservative estimate. It shows a significant but smaller positive relationship between feminist publications and women's subsequent graduation rates, as well as a small but significant positive relationship for the gender medicine paradigm. Again it shows no effect for the sex difference paradigm.

I cross validated Models 1 and 2 in order to test their robustness to the selection of fields and subfields by fitting the same model to subsets of the data with each field or subfield removed. This revealed highly consistent effect sizes (not shown) for all three paradigms no matter which subfield was removed in both Models 1 and 2. As expected Model 1's maximal specification did not always converge and encountered matrix algebra errors in some of the folds, making the more parsimonious specification of Model 2 more reliable (Bates et al. 2018). However, field level cross validation did reveal interesting variation indicating that sex research paradigms have different effects in psychology subfields than they do in others. Specifically, the effect of sex difference publications in a model without psychology subfields is significant and negative, while it is significant and positive in a model that includes only psychology subfields.

This is an indication that there is an interaction between psychology and the effect of sex difference research.

Model 4 tests that interaction by adding fixed effects for a dummy variable psychology and the interaction between that and sex difference research. In this model, the main effect of sex difference research, indicating its effect in life sciences outside of psychology, is significant and negative. While the coefficient for sex difference publications is smaller in absolute terms than the one for feminist biology publications (-4.79 for difference compared to 7.34 for feminism), it is larger in practical terms because sex difference publications are much more prevalent and have wider variance than feminist ones. A one standard deviation increase in the amount of difference publications in a field corresponds with a 2.00 point *decrease* in the future proportion of women graduates. A one standard deviation increase of feminist publications in Model 4 corresponds with a 1.15 point increase in women's representation.

The interaction between psychology and sex difference is large and significant, indicating that sex difference research has an unexpectedly *positive* effect on women's graduation rates in psychology subfields. The effect is small: -4.79 (main) + 7.44 (interaction) = 2.65 points more women graduates per point difference publications, or 1.11 per standard deviation. Log likelihood ratio tests indicate that Model 1 (maximal specification) fits better than Model 4 (interaction with psychology), which fits better than Model 2 (fixed, non interacted paradigm effects). Although the maximal model fits best, the additional specifications provide insight into the robustness of its results and into the dynamics of between-field variation.

### 4.5 Discussion and Conclusions

Using newly tabulated data from the National Science Foundation on the gender of PhD graduates in detailed research subfields since 1970 and a full copy of the Web of Science database, I showed that the content of sex research publications is related to the future gender ratio of life science PhDs. The publication of explicitly feminist research is associated with greater proportions of women earning PhDs in life science subfields in later years. The effects of research paradigms on PhD gender ratios can be context-dependent. In biological and health sciences, the prevalence of the sex difference paradigm has a negative relationship with the proportion of women subsequently earning PhDs. That effect is surprisingly reversed in psychology subfields.

I argued that scientific publications influence the gender ratio of PhDs by influencing and reflecting professional cultures. While the data in this paper cannot reveal specific causal mechanisms for that process, I review a wide range of literature that theorizes and tests such mechanisms. Lab experiments show that research can reinforce prejudice and stereotype threat through neurogenic essentialism, pushing girls out of STEM, while feminist approaches can buffer those effects (Moè 2012). I show the same time-ordered associations exist across the life sciences over a 5 decade span, suggesting that publications about sex influence the schemas of gender inequality in a field's professional culture. This is not surprising. Scientists generally believe their own field's research. Numerous sources point out the detrimental personal effects of research questioning the fundamental nature and merit of social groups on scientists from those groups. Such work makes it difficult to form the concordant professional identities that are key to success in graduate school (Costello 2005). Prior work has also shown that gender influences students' interest in research topics, which in turn influences selection into fields. The popularity of those topics within fields determines resource access and success in life science graduate programs. Further, my findings are generally consistent with paradigm practitioners' own agendas to either increase (feminist biology) or resist (sex difference) women's representation in science (Hubbard 1990).

My results highlight important, often neglected dimensions of other research on women in STEM. Studies of ideal worker norms and field-specific ability beliefs have focused on one half of the story: variation in the expectations of a good worker. They take for granted the (well documented) negative cultural association between "woman" and "genius," "objective," or "dedicated worker." But I demonstrate that such associations are not uniform across time or scientific fields. Because beliefs about women's nature vary from the professional culture of one field to another and one year to the next, future work on gendered occupational segregation would benefit from including explicit measures of gender beliefs alongside measures of work expectations. Such measures can be constructed for other academic fields or even occupations where practitioners do not regularly publish by using different data.

More generally, my finding that there is an association between the content of published research on sex and the future sex ratio of PhD graduates, which is robust across a wide range of life sciences and a long time span, drives home a well established but under-appreciated claim of feminist science studies. The content of science and scientists themselves are not separable as independent, objective entities. This point is all to often reduced to the claim that scientists' social positions influence their work and its reception. This is true. But it leaves key aspects of gender and science epistemology unexamined (Brickhouse 2001). I reverse the question to remind us that the way scientists talk about sex matters, too (Hubbard 1990; Martin 1991). As numerous scholars in feminist and queer methods have noted from their own experience, we are profoundly shaped and affected by the content of our research (Compton, Meadow, and Schilt 2018).

Although the relationship between research content and PhD gender is modest, efforts to understand women's under-representation in science are incomplete without consideration of the content of science itself. Publishing feminist research may be one way to increase women's representation. Yet not all approaches to feminist research seem to be equal. The critical feminist biology paradigm, which aims to complicate binaries and emphasize ongoing processes of becoming sexed and gendered has a positive effect on women entering scientific fields. The same does not appear to be true for gender medicine, which embraces notions of innate sex difference, albeit in the name of women's health. Perhaps gender medicine is ineffective because it breaks theoretically and politically with earlier feminist and women's health movements (for a critique to this effect, see Epstein 2007).

More than simply printing critical feminist research, we would do well to learn from the processes by which it relates to professional culture. Feminist biology, unlike gender medicine, is not a "just add women" recipe. It involves fundamentally rethinking received wisdom about the nature of men and women, questioning in our research and our lives how things come to be sexed and gendered and what ways that matters. It means thinking in ways that give life to, rather than pathologizing, intersex, trans, nonbinary, and queer people. It means letting go of homogenous notions of a "universal woman," whose character exists without race, class, nation, religion, history, let alone individuality. It means valuing, rewarding, and funding scientists who bring such perspectives to their departments and research. And it means championing those perspectives to junior scholars as exemplars of how professional scientists think. These are transformations of research agendas, yes. But they are also transformations of mind and culture.

My findings suggest several areas for additional study. The unexpected positive relationship between sex difference research and women's graduation in psychology raises important questions: does something about psychology training inoculate students against harmful stereotypes in a way other fields could learn from? Does research on sex differences carry a different focus or tone in psychology that makes it less hostile to women? Does psychology have more trans-exclusive radical feminists who see the paradigm as supporting their interests? My data are limited to sex research in the life sciences, but many other fields have feminist paradigms competing with more dominant approaches. In chemistry and physics scholars have made a compelling case that their fields can be fundamentally rethought as "liberatory science" using feminist theory; that teaching it this way would diversify the range of students in the field; and that feminist theory can even meaningfully inform core substance of physics and chemistry research (e.g. Barad 1995; Barton 1997; Prescod-Weinstein 2020). Further, there has long been fierce competition between paradigms in research on other human social categories—such as race, ability, and sexuality—in the life sciences (Epstein 2007; Panofsky 2014; Waidzunas 2015). Future work should construct measures of the competing paradigms in these domains and, where possible, connect it with demographic data in order to evaluate the generalizability of my findings to other domains.

## Chapter 5 Conclusion

Social conceptions of sex are part of a feedback cycle in science, wherein scientists' cultural assumptions about sex influence how they think about, design, interpret, and present research. The research they produce influences both their local disciplinary culture and the broader culture of the societies they live in, sometimes through the scientists' own public engagement efforts. Sex and science are both hugely important aspects of society, interesting in their own right and the subjects of substantial sociological subfields. As the chapters of this dissertation show, investigating aspects of the sex-science feedback cycle can also reveal insights in other areas. Chapter 2 demonstrates that the prevailing understanding of how algorithmic harm happens, and thus how it can be remedied, is misguided. Scientists' use of algorithms for sexing brains does things right: It pays close attention to measurement and theoretical implications and it tightly aligns models with theory. Nevertheless, it largely results in reinforcing the dominant, sexist cultural framing of sex and gender.

Chapter 3 turns from the doing of science to its presentation for both scientific and lay audiences. It demonstrates the importance of historical framing, particularly revisionist historical frames, in establishing the authority of claims about the biological nature of sex. That revisionism is about the history of sex research, to be sure, but as a chapter in an edited volume on *Far Right Revisionism and the End of History* (Valencia-García 2020), it situates the techniques of legitimacy and persuasion from sex science in a broader field of right wing political tactics. It might seem odd to situate sex science in this way, even the most essentialist forms of it, given that its proponents often describe themselves as "liberals" or "feminists." But such is the reality of contemporary right wing politics in Europe and North America, where Republicans and Tories have introduced hundreds of pieces of legislation to regulate and enforce a strict, biological binary understanding of sex in a moral panic lasting longer and reaching further than its twin moral panic over "critical race theory," and where far right governments in Hungary, Poland, and elsewhere have taken even more drastic stances against "gender ideology," including attacks on civil liberties and shuttering university departments. The biology of sex is, and has always been, a battle ground about women's (and men's) proper place in society. This is not secondary to or separable from the science; it is as blunt and integral in foundational texts like Charles Darwin's *Descent of Man* as racism is to sociology's own Emile Durkheim.

Chapter 4, about the feedback process whereby scientists' sex research is related to the future composition of the scientific workforce, is also a unique approach to examining processes of occupational segregation. With few exceptions (e.g. journalism, politics), most occupations do not produce regular, voluminous textual records of workers' beliefs. This limits analyses of the relationship between worker beliefs and occupational demographic composition to cross-sectional interviews and surveys. In the case of scientists writing about human nature and group differences, however, I was able to measure workforce beliefs across five decades using scientists' publications. In doing so, I demonstrated that the predominant focus in the occupational segregation literature on variation in job expectations is only part of the story. Variation in expectations about sex/gender matter, too. Put differently, jobs are thought of and structured in ways that put them more or less at odds with women, but also occupants of some jobs think of women in ways that put them more or less at odds with the job.

Sex is, of course, only one of many systems for classifying people that has been subject to debates about biologization of human groups. Race, too, has a long history of biologizing and essentializing differences between people, with fierce scientific and social conflicts. Notably, however, science on the biology of race and sex have taken radically different courses in the last decade. The National Institutes of Health mandate the use of "sex as a biological variable" in all funded research (NIH 2015), and numerous journals and professional societies have sprung up to advance this agenda (Epstein 2007; Pape 2021). Simultaneously, organs of the NIH and other professional scientific societies insist that race is *not* biological (Umek and Fischer 2020), some biology journals have begun to ban the word "race" on the grounds that it is not a valid scientific concept (Trujillo et al. 2021), and whole fields in the life sciences assiduously avoid discussion of race (Rollins 2021). While race is certainly not gone from the biological sciences that claim to have abandoned it (e.g. Morning 2007), and an isolated handful continue to openly work on race science (Panofsky 2018), the stark rhetorical contrast between how scientists discuss the legitimacy of race and sex as biological categories is inescapable.

This is even more stark compared with early scientific articulations that treated race and sex as inextricably intermingled, with claims such as that women "represent the most inferior forms of human evolution and that they are closer to children and savages than to an adult, civilized man" (Gould [1981] 1996 quoting and translating; Le Bon 1879:60–61). Indeed, the scientific projects of sexing and racing bodies arose together, co-producing one another as part of the larger colonial project of domination and control (Amadiume 1987; Mahan 2017; Patil 2018; Scheuerman et al. 2021).

This is the puzzle to which I will turn my attentions after this dissertation. How did it come to be that our scientific efforts toward making up people—to use Ian Hacking's term—have diverged so sharply for social categories like sex and race? What are the conditions, both in the substance of the scientific work and in the cultural/structural conditions of that work guiding the paths of sex and race science? There is no doubt that both are important to the story. Many of the characteristics used to group people by race differ in how they operate, biologically, from the characteristics used to group people by sex. In some senses, sex is more biologically fixed than race: children of different sex parents tend to be one sex with little ambiguity, not an intermediate combination of sexes in the way that has long troubled scientists interested in racial classification. Yet in other ways sex is less biologically fixed than race: Testosterone and estrogen supplements are common, safe, and efficacious means of altering bodies, while injecting melanotan is carcinogenic, priapigenic, and only minimally effective at changing skin color (Dreyer, Amer, and Fraser 2019). Simultaneously, the social and political conditions for scientific claims about sexual and racial difference have not been the same. Popular writing about race, intelligence, and genetics in the 1990s such as *The Bell Curve* sparked much more heated debate and criticism than contemporary work on sex, including *Men are from Mars, Women are from Venus*.

It is my hope that this dissertation, and the work I do after it, will help shed light on the scientific construction of human groups and their differences, not simply for the sake of generating knowledge or advancing my academic career, but in order that we might remember that what is made can always be made differently. While no one has the magical power to instantly remake our world by fiat, understanding first that aspects of our world like sex and race are constructed, and second how those constructions have been produced, challenged, and changed, opens the door for conscious efforts to build a better world.

## References


Abbott, Andrew. 1988. *The System of Professions*. Chicago: University of Chicago Press.

Adebayo, Julius, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. 2018. "Sanity Checks for Saliency Maps." P. 11 in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*. Montreal.

Adeli, Ehsan, Qingyu Zhao, Natalie M. Zahr, Aimee Goldstone, Adolf Pfefferbaum, Edith V. Sullivan, and Kilian M. Pohl. 2020. "Deep Learning Identifies Morphological Determinants of Sex Differences in the Pre-Adolescent Brain." *NeuroImage* 223:117293. doi: 10.1016/j.neuroimage.2020.117293.

Alexander, Gerianne M., and Melissa Hines. 2002. "Sex Differences in Response to Children's Toys in Nonhuman Primates (Cercopithecus Aethiops Sabaeus)." *Evolution and Human Behavior* 23:467–79.

Almeling, Rene. 2020. *GUYnecology: The Missing Science of Men's Reproductive Health*. Oakland: University of California Press.

Amadiume, Ifi. 1987. "Colonialism and the Erosion of Women's Power." in *Male Daughters, Female Husbands: Gender and Sex in African Society*. London: Zed Books.

Anderson, Michael L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. MIT Press.

Anderson, Nathaniel E., Keith A. Harenski, Carla L. Harenski, Michael R. Koenigs, Jean Decety, Vince D. Calhoun, and Kent A. Kiehl. 2019. "Machine Learning of Brain Gray Matter Differentiates Sex in a Large Forensic Sample." *Human Brain Mapping* 40(5):1496–1506. doi: https://doi.org/10.1002/hbm.24462.

Arribas-Ayllon, Michael, Andrew Bartlett, and Katie Featherstone. 2010. "Complexity and Accountability: The Witches' Brew of Psychiatric Genetics." *Social Studies of Science* 40(4):499–524. doi: 10.1177/0306312710363511.

Arslan, Salim, Sofia Ira Ktena, Ben Glocker, and Daniel Rueckert. 2018. "Graph Saliency Maps Through Spectral Convolutional Networks: Application to Sex Classification with Brain Connectivity." Pp. 3–13 in *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*, *Lecture Notes in Computer Science*, edited by D. Stoyanov, Z. Taylor, E. Ferrante, A. V. Dalca, A. Martel, L. Maier-Hein, S. Parisot, A. Sotiras, B. Papiez, M. R. Sabuncu, and L. Shen. Cham: Springer International Publishing.

Azari, Abigail R., Jeffrey W. Lockhart, Michael W. Liemohn, and Xianzhe Jia. 2020. "Incorporating Physical Knowledge Into Machine Learning for Planetary Space Physics." *Frontiers in Astronomy and Space Sciences* 7:36. doi: 10.3389/fspas.2020.00036.

Bailey, Kimberlyn A., David Horacek, Steven Worthington, Ampalavanar Nanthakumar, Scott Preston, and Carolina C. Ilie. 2019. "STEM/Non-STEM Divide Structures Undergraduate Beliefs About Gender and Talent in Academia." *Frontiers in Sociology* 4. doi: 10.3389/fsoc.2019.00026.

Baldinger-Melich, Pia, Maria F. Urquijo Castro, René Seiger, Anne Ruef, Dominic B. Dwyer, Georg S. Kranz, Manfred Klöbl, Joseph Kambeitz, Ulrike Kaufmann, Christian Windischberger, Siegfried Kasper, Peter Falkai, Rupert Lanzenberger, and Nikolaos Koutsouleris. 2020. "Sex Matters: A Multivariate Pattern Analysis of Sex- and Gender-Related Neuroanatomical Differences in Cis- and Transgender Individuals Using Structural Magnetic Resonance Imaging." *Cerebral Cortex* 30(3):1345–56. doi: 10.1093/cercor/bhz170.

Bansal, Naman, Chirag Agarwal, and Anh Nguyen. 2020. "SAM: The Sensitivity of Attribution Methods to Hyperparameters." Pp. 8673–83 in.

Bao, Ai-Min, and Dick F. Swaab. 2011. "Sexual Differentiation of the Human Brain: Relation to Gender Identity, Sexual Orientation and Neuropsychiatric Disorders." *Frontiers in Neuroendocrinology* 32(2):214–26.

Barad, K. 1995. "A Feminist Approach to Teaching Quantum Physics." Pp. 43–75 in *Teaching the majority*, edited by Rosser, Sue V. Teachers College Press New York.

Baron-Cohen, Simon. 1999. "The Extreme Male-Brain Theory of Autism." Pp. 401–29 in *Neurodevelopmental Disorders*, edited by H. Tager-Flusberg. MIT Press.

Baron-Cohen, Simon. 2003. *The Essential Difference: Male and Female Brains and the Truth about Autism*. New York: Basic Books.

Baron-Cohen, Simon. 2007. "Does Biology Play Any Role in Sex Differences in the Mind?" Pp. 77–97 in *The Future of Gender*, edited by J. Browne. Cambridge: Cambridge University Press.

Baron-Cohen, Simon. 2019. "The Gendered Brain by Gina Rippon Review — Do Men and Women Have Different Brains?" *The Times*, March 8.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68(3):255–78. doi: 10.1016/j.jml.2012.11.001.

Barton, Angela Calabrese. 1997. "Liberatory Science Education: Weaving Connections Between Feminist Theory and Science Education." *Curriculum Inquiry* 27(2):141–63. doi: 10.1080/03626784.1997.11075486.

Bashford, Alison, and Philippa Levine. 2010. *The Oxford Handbook of the History of Eugenics*. Oxford University Press, USA.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2018. "Parsimonious Mixed Models." *ArXiv:1506.04967 [Stat]*.

Becker, Howard Saul. [1963] 1973. *Outsiders; Studies in the Sociology of Deviance*. London: Free Press of Glencoe.

Belavy, Daniel L., Patrick J. Owen, and Patricia M. Livingston. 2020. "Do Successful PhD Outcomes Reflect the Research Environment Rather than Academic Ability?" *PLOS ONE* 15(8). doi: 10.1371/journal.pone.0236327.

Benjamin, Ruha. 2016. "Catching Our Breath: Critical Race STS and the Carceral Imagination." *Engaging Science, Technology, and Society* 2:145–56. doi: 10.17351/ests2016.70.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. 1 edition. Medford, MA: Polity.

Birhane, Abeba. 2021. "The Impossibility of Automating Ambiguity." *Artificial Life* 27(1):44–61. doi: 10.1162/artl_a_00336.

Bleier, Ruth. 1986. "Sex Differences Research: Science or Belief?" Pp. 147–64 in *Feminist Approaches to Science*. New York: Pergamon.

Bleier, Ruth. 1997. *Science and Gender: A Critique of Biology and Its Theories on Women*. New York: Teachers College Press.

Bluhm, Robyn. 2012. "Beyond Neurosexism: Is It Possible to Defend the Female Brain?" Pp. 230–45 in *Neurofeminism: Issues at the Intersection of Feminist Theory and Cognitive Science*, edited by R. Bluhm, A. J. Jacobson, and H. L. Maibom. London: Palgrave Macmillan UK.

Bluhm, Robyn. 2021. "Neurosexism and Our Understanding of Sex Differences in the Brain." in *Handbook to Feminist Philosophy of Science*, edited by K. Intemann and S. Crasnow. New York: Routledge.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Cambridge: Harvard University Press.

Bourdieu, Pierre. 1985. "The Social Space and the Genesis of Groups." *Theory and Society* 14(6):723–44.

Bowker, Geoffrey C. 2014. "Big Data, Big Questions: The Theory/Data Thing." 5.

Bowker, Geoffrey C, and Susan Leigh Star. 1999. "Some Tricks of the Trade in Analyzing Classification." Pp. 33–50 in *Sorting Things Out: Classification and Its Concequences*. Cambridge: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things out: Classification and Its Consequences*. Cambridge, Mass: MIT Press.

Brennan, Daniel, Tingting Wu, and Jin Fan. 2021. "Morphometrical Brain Markers of Sex Difference." *Cerebral Cortex*. doi: 10.1093/cercor/bhab037.

Brescoll, Victoria, and Marianne LaFrance. 2004. “The Correlates and Consequences of Newspaper Reports of Research on Sex Differences.” *Psychological Science* 15(8):515–20. doi: 10.1111/j.0956-7976.2004.00712.x.

Brickhouse, Nancy W. 2001. “Embodying Science: A Feminist Perspective on Learning.” *Journal of Research in Science Teaching* 38(3):282–95. doi: 10.1002/1098-2736(200103)38:3<282::AID-TEA1006>3.0.CO;2-0.

Bucher, Taina. 2016. “Neither Black Nor Box: Ways of Knowing Algorithms.” Pp. 81–98 in *Innovative Methods in Media and Communication Research*, edited by S. Kubitschko and A. Kaun. Cham: Springer International Publishing.

Buffington, Catherine, Benjamin Cerf, Christina Jones, and Bruce A. Weinberg. 2016. “STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census.” *American Economic Review* 106(5):333–38. doi: 10.1257/aer.p20161124.

Burrell, Jenna, and Marion Fourcade. 2021. “The Society of Algorithms.” *Annual Review of Sociology* 47(1):annurev-soc-090820-020800. doi: 10.1146/annurev-soc-090820-020800.

Bzdok, Danilo. 2017. “Classical Statistics and Statistical Learning in Imaging Neuroscience.” *Frontiers in Neuroscience* 11. doi: 10.3389/fnins.2017.00543.

Cahill, Larry. 2010. “Sex Influences on Brain and Emotional Memory: The Burden of Proof Has Shifted.” Pp. 29–40 in *Progress in Brain Research*. Vol. 186, *Sex Differences in the Human Brain, their Underpinnings and Implications*, edited by I. Savic. Elsevier.

Cahill, Larry. 2014. “Fundamental Sex Difference in Human Brain Architecture.” *Proceedings of the National Academy of Sciences* 111(2):577–78. doi: 10.1073/pnas.1320954111.

Cahill, Larry. 2017. “An Issue Whose Time Has Come.” *Journal of Neuroscience Research* 95(1–2):12–13.

Cahill, Larry. 2019. “Denying the Neuroscience of Sex Differences.” *Quillette*. Retrieved March 30, 2019 (https://quillette.com/2019/03/29/denying-the-neuroscience-of-sex-differences/).

Campolo, Alexander, and Kate Crawford. 2020. “Enchanted Determinism: Power without Responsibility in Artificial Intelligence.” *Engaging Science, Technology, and Society* 6:1–19. doi: 10.17351/ests2020.277.

Carlson, Thomas, Erin Goddard, David M. Kaplan, Colin Klein, and J. Brendan Ritchie. 2018. “Ghosts in Machine Learning for Cognitive Neuroscience: Moving from Data to Theory.” *NeuroImage* 180:88–100. doi: 10.1016/j.neuroimage.2017.08.019.

Cech, Erin A. 2013a. “Ideological Wage Inequalities? The Technical/Social Dualism and the Gender Wage Gap in Engineering.” *Social Forces* 91(4):1147–82. doi: 10.1093/sf/sot024.

Cech, Erin A. 2013b. “The (Mis)Framing of Social Justice: Why Ideologies of Depoliticization and Meritocracy Hinder Engineers’ Ability to Think About Social Injustices.” Pp. 67–84 in *Engineering Education for Social Justice*. Vol. 10, edited by J. Lucena. Dordrecht: Springer Netherlands.

Cech, Erin A. 2014. "Culture of Disengagement in Engineering Education?" *Science, Technology, & Human Values* 39(1):42–72. doi: 10.1177/0162243913504305.

Cech, Erin A. 2017. "Rugged Meritocratists: The Role of Overt Bias and the Meritocratic Ideology in Trump Supporters' Opposition to Social Justice Efforts." *Socius* 3:2378023117712395. doi: 10.1177/2378023117712395.

Cech, Erin A. 2021. *The Trouble with Passion: How Searching for Fulfillment at Work Fosters Inequality*. Oakland: Univ of California Press.

Cech, Erin A., and Mary Blair-Loy. 2019. "The Changing Career Trajectories of New Parents in STEM." *Proceedings of the National Academy of Sciences* 116(10):4182–87. doi: 10.1073/pnas.1810862116.

Cech, Erin A., Mary Blair-Loy, and Laura E. Rogers. 2018. "Recognizing Chilliness: How Schemas of Inequality Shape Views of Culture and Climate in Work Environments." *American Journal of Cultural Sociology* 6(1):125–60. doi: 10.1057/s41290-016-0019-1.

Chekroud, Adam M., Emily J. Ward, Monica D. Rosenberg, and Avram J. Holmes. 2016. "Patterns in the Human Brain Mosaic Discriminate Males from Females." *Proceedings of the National Academy of Sciences* 113(14):E1968–E1968. doi: 10.1073/pnas.1523888113.

Chen, Chaomei, Zhigang Hu, Jared Milbank, and Timothy Schultz. 2013. "A Visual Analytic Study of Retracted Articles in Scientific Literature." *Journal of the American Society for Information Science and Technology* 64(2):234–53. doi: 10.1002/asi.22755.

Chen, Chen, Xuyu Cao, and Lixia Tian. 2019. "Partial Least Squares Regression Performs Well in MRI-Based Individualized Estimations." *Frontiers in Neuroscience* 13. doi: 10.3389/fnins.2019.01282.

Cheney-Lippold, John. 2017. *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.

Ching, Boby Ho-Hong, and Jason Teng Xu. 2018. "The Effects of Gender Neuroessentialism on Transprejudice: An Experimental Study." *Sex Roles* 78(3):228–41. doi: 10.1007/s11199-017-0786-3.

Cicourel, Aaron Victor. 1964. *Method and Measurement in Sociology*. [New York].

Colapinto, John. 2000. *As Nature Made Him: The Boy Who Was Raised as a Girl*. 1st ed. New York: Harper Collins.

Coleman, Jill M., and Ying-Yi Hong. 2008. "Beyond Nature and Nurture: The Influence of Lay Gender Theories on Self-Stereotyping." *Self and Identity* 7(1):34–53. doi: 10.1080/15298860600980185.

Collins, Patricia Hill. [1990] 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd ed. New York: Routledge.

Compton, D'Lane, Tey Meadow, and Kristen Schilt, eds. 2018. *Other, Please Specify*. Oakland: University of California Press.

Costello, Carrie Yang. 2005. *Professional Identity Crisis: Race, Class, Gender, and Success at Professional Schools*. Nashville: Vanderbilt University Press.

Crawford, Kate, and Trevor Paglen. 2019. *Excavating AI: The Politics of Training Sets for Machine Learning*. New York: The AI Now Institute, NYU.

Darwin, Charles. [1871] 1901. *Descent of Man, and Selection in Relation to Sex*. D. Appleton.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. Princeton: Princeton University Press.

David, Sean P., Florian Naudet, Jennifer Laude, Joaquim Radua, Paolo Fusar-Poli, Isabella Chu, Marcia L. Stefanick, and John P. A. Ioannidis. 2018. "Potential Reporting Bias in Neuroimaging Studies of Sex Differences." *Scientific Reports* 8(1):6082. doi: 10.1038/s41598-018-23976-1.

Davis, Georgiann. 2015. *Contesting Intersex: The Dubious Diagnosis*. New York: NYU Press.

Del Giudice, Marco. 2021. "Ideological Bias in the Psychology of Sex and Gender." in *Political Bias in Psychology: Nature, Scope, and Solutions*, edited by C. Frisby, W. O'Donohue, R. Redding, and S. Lilienfeld. Springer.

Del Giudice, Marco, Tom Booth, and Paul Irwing. 2012. "The Distance Between Mars and Venus: Measuring Global Sex Differences in Personality." *PLOS ONE* 7(1):e29265. doi: 10.1371/journal.pone.0029265.

Del Giudice, Marco, Richard Lippa, David Puts, Drew Bailey, J. Michael Bailey, and David Schmitt. 2015. "Mosaic Brains? A Methodological Critique of Joel et al. (2015)."

Dhamala, Elvisha, Keith W. Jamison, Mert R. Sabuncu, and Amy Kuceyeski. 2020. "Sex Classification Using Long-Range Temporal Dependence of Resting-State Functional MRI Time Series." *Human Brain Mapping* 41(13):3567–79. doi: https://doi.org/10.1002/hbm.25030.

Di Bitetti, Mario S., and Julián A. Ferreras. 2017. "Publish (in English) or Perish: The Effect on Citation Rate of Using Languages Other than English in Scientific Publications." *Ambio* 46(1):121–27. doi: 10.1007/s13280-016-0820-7.

Donovan, Brian M., Molly A. M. Stuhlsatz, Daniel C. Edelson, and Zoë E. Buck Bracey. 2019. "Gendered Genetics: How Reading about the Genetic Basis of Sex Differences in Biology Textbooks Could Affect Beliefs Associated with Science Gender Disparities." *Science Education* 103(4):719–49. doi: 10.1002/sce.21502.

Dreger, Alice Domurat. 2015. *Galileo's Middle Finger: Heretics, Activists, and the Search for Justice in Science*. New York: Penguin Press.

Dreyer, Barend Albert, Tarik Amer, and Michael Fraser. 2019. "Melanotan-Induced Priapism: A Hard-Earned Tan." *BMJ Case Reports* 12(2):e227644. doi: 10.1136/bcr-2018-227644.

Duarte-Carvajalino, Julio M., Neda Jahanshad, Christophe Lenglet, Katie L. McMahon, Greig I. de Zubicaray, Nicholas G. Martin, Margaret J. Wright, Paul M. Thompson, and Guillermo Sapiro. 2012. "Hierarchical Topological Network Analysis of Anatomical Human Brain Connectivity and Differences Related to Sex and Kinship." *NeuroImage* 59(4):3784–3804. doi: 10.1016/j.neuroimage.2011.10.096.

Duggan, Lisa. 2003. *Twilight of Equality?: Neoliberalism, Cultural Politics, and the Attack on Democracy*. Boston: Beacon Press.

Dumit, Joseph. 2014. "How (Not) to Do Things with Brain Images." Pp. 291–314 in *Representation in Scientific Practice Revisited*, edited by C. Coopman, J. Vertesi, M. Lynch, and S. Woolgar. The MIT Press.

Durkheim, Emile, and Marcel Mauss. [1903] 1967. *Primitive Classification*. Chicago, IL: University of Chicago Press.

Ecker, Christine. 2019. "Notice of Retraction and Replacement: Ecker et al. Association between the Probability of Autism Spectrum Disorder and Normative Sex-Related Phenotypic Diversity in Brain Structure. JAMA Psychiatry. 2017;74(4):329-338." *JAMA Psychiatry* 76(5):549–50. doi: 10.1001/jamapsychiatry.2018.4296.

Eckes, Thomas, and Hanns M. Trautner. 2012. *The Developmental Social Psychology of Gender*. Psychology Press.

Eliot, Lise. 2009. *Pink Brain, Blue Brain: How Small Differences Grow into Troublesome Gaps - and What We Can Do about It*. Boston: Mariner.

Eliot, Lise, Adnan Ahmed, Hiba Khan, and Julie Patel. 2021. "Dump the 'Dimorphism': Comprehensive Synthesis of Human Brain Studies Reveals Few Male-Female Differences beyond Size." *Neuroscience & Biobehavioral Reviews* 125:667–97. doi: 10.1016/j.neubiorev.2021.02.026.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 2011. "Processing Fieldnotes: Coding and Meaning." in *Writing Ethnographic Fieldnotes*. Chicago: University of Chicago Press.

Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. Berkeley: University of California Press.

Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Edition. New York, NY: St. Martin's Press.

Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42(May):1–30. doi: 10.1146/annurev-soc-081715-074206.

Fausto-Sterling, Anne. 1992a. "Building Two-Way Streets: The Case of Feminism and Science." *NWSA Journal* 4(3):336–49.

Fausto-Sterling, Anne. 1992b. *Myths of Gender: Biological Theories about Women and Men*. 2nd ed. New York: Basic Books.

Fausto-Sterling, Anne. 2000. *Sexing the Body: Gender Politics and the Construction of Sexuality*. Basic Books.

Fausto-Sterling, Anne. 2005. "The Bare Bones of Sex: Part 1-Sex and Gender." *Signs* 30(2):1491–1527.

Fehr, Tristan. 2020. "Essentially a Lie: Challenging Biological Essentialist Interpretations of Transgender Neurology." Pp. 387–411 in *Far-Right Revisionism and the End of History: Alt/Histories*, edited by L. D. Valencia-García. New York : Routledge Taylor & Francis Group, [2020] |: Routledge.

Feis, Delia-Lisa, Kay H. Brodersen, D. Yves von Cramon, Eileen Luders, and Marc Tittgemeyer. 2013. "Decoding Gender Dimorphism of the Human Brain Using Multimodal Anatomical and Diffusion MRI Data." *NeuroImage* 70:250–57. doi: 10.1016/j.neuroimage.2012.12.068.

Ferber, Alona, and Judith Butler. 2020. "Judith Butler on the Culture Wars, JK Rowling and Living in 'Anti-Intellectual Times.'" *New Statesman*, September 22.

Field, Taylor L., and Greggor Mattson. 2016. "Parenting Transgender Children in PFLAG." *Journal of GLBT Family Studies* 1–17.

Fine, Cordelia. 2010. *Delusions of Gender: The Real Science Behind Sex Differences*. London.

Fine, Cordelia. 2013. "Is There Neurosexism in Functional Neuroimaging Investigations of Sex Differences?" *Neuroethics* 6(2):369–409. doi: 10.1007/s12152-012-9169-1.

Fitsch, Hannah. 2021. "Reflections on Binary Sex/Gender Categorization in Magnetic Resonance Tomography and Its Future Challenges." *Frontiers in Sociology* 6.

Forbes, TehQuin D. 2020. "Queer-Free Majors?: LGBTQ + College Students' Accounts of Chilly and Warm Academic Disciplines." *Journal of LGBT Youth*. doi: 10.1080/19361653.2020.1813673.

Foucault, Michel. 1990. *The History of Sexuality: Volume 1: An Introduction*. New York: Vintage Books.

Fourcade, Marion, and Kieran Healy. 2016. "Seeing like a Market." *Socio-Economic Review* mww033. doi: 10.1093/ser/mww033.

Fujimura, Joan H. 2006. "Sex Genes: A Critical Sociomaterial Approach to the Politics and Molecular Genetics of Sex Determination." *Signs: Journal of Women in Culture and Society* 32(1):49–82.

Gansen, Heidi M., and Karin A. Martin. 2018. "Becoming Gendered." Pp. 83–93 in *Handbook of the Sociology of Gender*, *Handbooks of Sociology and Social Research*, edited by B. J. Risman, C. M. Froyum, and W. J. Scarborough. Cham: Springer International Publishing.

Garfinkel, Harold. [1967] 2006. “Passing and the Managed Achievement of Sex Status in an ‘Intersexed’ Person.” Pp. 58–83 in *The Transgender Studies Reader*, edited by S. Stryker and S. Whittle. New York: Routledge.

Geary, David C. 2010. *Male, Female: The Evolution of Human Sex Differences*. Second Edition. Washington DC: American Psycological Association.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. “Datasheets for Datasets.” *ArXiv:1803.09010 [Cs]*.

Gelman, Andrew, Greggor Mattson, and Daniel Simpson. 2018. “Gaydar and the Fallacy of Decontextualized Measurement.” *Sociological Science* 5:270–80. doi: 10.15195/v5.a12.

Glocker, Ben, Robert Robinson, Daniel C. Castro, Qi Dou, and Ender Konukoglu. 2019. “Machine Learning with Multi-Site Imaging Data: An Empirical Study on the Impact of Scanner Effects.” *ArXiv:1910.04597 [Cs, Eess, q-Bio]*.

Goffman, Erving. [1963] 1986. *Stigma; Notes on the Management of Spoiled Identity*. Touchstone. New York: Simon & Schuster.

Górriz, Juan M., Javier Ramírez, F. Segovia, Francisco J. Martínez, Meng-Chuan Lai, Michael V. Lombardo, Simon Baron-Cohen, and John Suckling. 2018. “A Machine Learning Approach to Reveal the NeuroPhenotypes of Autisms.” *International Journal of Neural Systems* 29(07):1850058. doi: 10.1142/S0129065718500582.

Gould, Stephen Jay. [1981] 1996. *The Mismeasure of Man*. Rev. and expanded. New York: Norton.

Gray, John. 1997. *Mars and Venus on a Date: A Guide for Navigating the 5 Stages of Dating to Create a Loving and Lasting Relationship*. 1st ed. New York, NY: Harper Collins.

Greenblatt, David, Jerold Harmatz, and Thomas Roth. 2019. “Zolpidem and Gender: Are Women Really At Risk?” *Journal of Clinical Psychopharmacology* 39(3):189–99.

Gross, Paul R., and Norman Levitt. 1994. *Higher Superstition: The Academic Left and Its Quarrels with Science*. Baltimore: Johns Hopkins University Press.

Gur, Ruben C., and Raquel E. Gur. 2017. “Complementarity of Sex Differences in Brain and Behavior: From Laterality to Multimodal Neuroimaging.” *Journal of Neuroscience Research* 95(1–2):189–99.

Hacker, Helen Mayer. 1953. “Marx, Weber and Pareto on the Changing Status of Women.” *The American Journal of Economics and Sociology* 12(2):149–62.

Hacking, Ian. 1986. “Making Up People.” Pp. 222–36 in *Reconstructing Individualism: Autonomy, Individuality and the Self in Western Thought*, edited by T. C. Heller, M. Sosna, and D. E. Wellbery. Stanford: Stanford University Press.

Hacking, Ian. 2007. “Kinds of People: Moving Targets.” Pp. 285–317 in *Proceedings of the British Academy*. Vol. 151. British Academy.

Hall, Diana Long. 1976. "Biology, Sex Hormones and Sexism in the 1920s." Pp. 81–96 in *Women and philosophy: toward a theory of liberation*, edited by C. C. Gould and M. W. Wartofsky. Perigee.

Halpern, Diane F. 2012. *Sex Differences in Cognitive Abilities*. 4th Edition. New York: Psychology Press.

Hancox-Li, Leif, and I. Elizabeth Kumar. 2021. "Epistemic Values in Feature Importance Methods: Lessons from Feminist Epistemology." Pp. 817–26 in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Haraway, Donna Jeanne. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14(3):575–99.

Harding, Sandra. 1993. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?" Pp. 49–82 in *Feminist Epistemologies*, edited by L. Alcoff and E. Potter. New York: Routledge.

Hassett, Janice M., Erin R. Siebert, and Kim Wallen. 2008. "Sex Differences in Rhesus Monkey Toy Preferences Parallel Those of Children." *Hormones and Behavior* 54(3):359–64.

Haxby, James. V. 2012. "Multivariate Pattern Analysis of FMRI: The Early Beginnings." *Neuroimage* 62(2):852–55. doi: 10.1016/j.neuroimage.2012.03.016.

Herschberger, Ruth. 1948. *Adam's Rib*. New York: Pellegrini & Cudahy.

Hess, Beth B. 1990. "Beyond Dichotomy: Drawing Distinctions and Embracing Differences." *Sociological Forum* 5(1):75–93.

Hilton, Emma, and Jenny Whyte. n.d. "Project Nettie." *Project Nettie*. Retrieved January 8, 2020 (https://projectnettie.wordpress.com/).

Hines, Melissa. 2004. *Brain Gender*. New York: Oxford University Press.

Hines, Melissa. 2011. "Gender Development and the Human Brain." *Annual Review of Neuroscience* 34(1):69–88.

Hines, Melissa, and Gerianne M. Alexander. 2008. "Commentary: Monkeys, Girls, Boys and Toys: A Confirmation Comment on 'Sex Differences in Toy Preferences: Striking Parallels between Monkeys and Humans.'" *Hormones and Behavior* 54(3):478–81.

Hird, Myra J. 2004a. "Feminist Matters: New Materialist Considerations of Sexual Difference." *Feminist Theory* 5(2):223–32. doi: 10.1177/1464700104045411.

Hird, Myra J. 2004b. "Naturally Queer." *Feminist Theory* 5(1):85–89. doi: 10.1177/1464700104040817.

Hochschild, Arlie Russell, and Anne Machung. 1989. *The Second Shift: Working Parents and the Revolution at Home*. New York: Viking.

Hoekzema, Elseline, Sebastian E. E. Schagen, Baudewijntje P. C. Kreukels, Dick J. Veltman, Peggy T. Cohen-Kettenis, Henriette Delemarre-van de Waal, and Julie Bakker. 2015.

"Regional Volumes and Spatial Volumetric Distribution of Gray Matter in the Gender Dysphoric Brain." *Psychoneuroendocrinology* 55:59–71. doi: 10.1016/j.psyneuen.2015.01.016.

Hubbard, Ruth. 1982. "Have Only Men Evolved?" Pp. 17–46 in *Biological Woman--The Convienent Myth*, edited by R. Hubbard, M. S. Henifin, and B. Fried. Cambridge, MA: Schenkman.

Hubbard, Ruth. 1990. *The Politics of Women's Biology*. New Brunswick: Rutgers University Press.

Hubbard, Ruth, Sandra Harding, Nancy Tuana, Sue V. Rosser, and Anne Fausto-Sterling. 1993. "Comments on Anne Fausto-Sterling's 'Building Two-Way Streets' [with Response]." *NWSA Journal* 5(1):45–81.

Huf, Wolfgang, Klaudius Kalcher, Roland N. Boubela, Georg Rath, Andreas Vecsei, Peter Filzmoser, and Ewald Moser. 2014. "On the Generalizability of Resting-State FMRI Machine Learning Classifiers." *Frontiers in Human Neuroscience* 8. doi: 10.3389/fnhum.2014.00502.

Hyde, Janet Shibley. 2005. "The Gender Similarities Hypothesis." *American Psychologist* 60(6):581–92.

Ichikawa, Kelsey. 2021. "The Trouble with Brain Scans." *Nautilus*, March 31.

Igo, Sarah Elizabeth. 2007. *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Harvard University Press.

Jackson, Virginia, Daniel Rosenberg, Travis D. Williams, Kevin R. Brine, Mary Poovey, Matthew Stanley, Ellen Gruber Garvey, Markus Krajewski PhD, Rita Raley, David Ribes, Steven J. Jackson, and Geoffrey C. Bowker. 2013. *"Raw Data" Is an Oxymoron*. edited by L. Gitelman. Cambridge, Massachusetts ; London, England: The MIT Press.

Jacobs, Abigail Z. 2021. "Measurement as Governance in and for Responsible AI." *ArXiv:2109.05658 [Cs]*.

Jacobs, Abigail Z., and Hanna Wallach. 2021. "Measurement and Fairness." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 375–85. doi: 10.1145/3442188.3445901.

Joel, Daphna, Zohar Berman, Ido Tavor, Nadav Wexler, Olga Gaber, Yaniv Stein, Nisan Shefi, Jared Pool, Sebastian Urchs, Daniel S. Margulies, Franziskus Liem, Jürgen Hänggi, Lutz Jäncke, and Yaniv Assaf. 2015. "Sex beyond the Genitalia: The Human Brain Mosaic." *Proceedings of the National Academy of Sciences* 112(50):15468–73.

Joel, Daphna, Ariel Persico, Moshe Salhov, Zohar Berman, Sabine Oligschläger, Isaac Meilijson, and Amir Averbuch. 2018. "Analysis of Human Brain Structure Reveals That the Brain 'Types' Typical of Males Are Also Typical of Females, and Vice Versa." *Frontiers in Human Neuroscience* 12. doi: 10.3389/fnhum.2018.00399.

Johnson, Matthew. 2020. *Undermining Racial Justice: How One University Embraced Inclusion and Inequality*. Ithaca [New York]: Cornell University Press.

Jordan-Young, Rebecca M. 2010. *Brain Storm: The Flaws in the Science of Sex Differences*. Cambridge: Harvard University Press.

Jordan-Young, Rebecca M., and Katrina Karkazis. 2019. *Testosterone: An Unauthorised Biography*. Cambridge, MA: Harvard University Press.

Jordan-Young, Rebecca M., and Raffaella I. Rumiati. 2012. "Hardwired for Sexism? Approaches to Sex/Gender in Neuroscience." *Neuroethics* 5(3):305–15.

Kane, Emily W. 2006. "'No Way My Boys Are Going to Be Like That!': Parents' Responses to Children's Gender Nonconformity." *Gender & Society* 20(2):149–76.

Karkazis, Katrina, and Rebecca M. Jordan-Young. 2015. "Debating a Testosterone 'Sex Gap.'" *Science* 348(6237):858–60. doi: 10.1126/science.aab1057.

Keyes, Os. 2018. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW):1–22. doi: 10.1145/3274357.

Keyes, Os, Zoë Hitzig, and Mwenza Blell. 2021. "Truth from the Machine: Artificial Intelligence and the Materialization of Identity." *Interdisciplinary Science Reviews* 46(1–2):158–75. doi: 10.1080/03080188.2020.1840224.

King, Ross D., Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Whelan, and Amanda Clare. 2009. "The Automation of Science." *Science* 324(5923):85–89. doi: 10.1126/science.1165620.

Kiviat, Barbara. 2019. "The Moral Limits of Predictive Practices: The Case of Credit-Based Insurance Scores." *American Sociological Review* 84(6):1134–58. doi: 10.1177/0003122419884917.

Klein, Viola. 1946. *The Feminine Character: History of an Ideology*. London: K. Paul, Trench, Trubner.

Kosinski, Michal. 2021. "Facial Recognition Technology Can Expose Political Orientation from Naturalistic Facial Images." *Scientific Reports* 11(1):100. doi: 10.1038/s41598-020-79310-1.

Lagos, Danya. 2019. "Hearing Gender: Voice-Based Gender Classification Processes and Transgender Health Inequality." *American Sociological Review* 84(5):801–27. doi: 10.1177/0003122419872504.

Lamont, Michèle. 2009. *How Professors Think*. Cambridge: Harvard University Press.

Lao, Zhiqiang, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M. Resnick, and Christos Davatzikos. 2004. "Morphological Classification of Brains via High-Dimensional Shape Transformations and Machine Learning Methods." *NeuroImage* 21(1):46–57. doi: 10.1016/j.neuroimage.2003.09.027.

Laqueur, Thomas. 2001. *Making Sex: Body and Gender from the Greeks to Freud*. Cambridge: Harvard University Press.

Laqueur, Thomas Walter. 1990. *Making Sex: Body and Gender from the Greeks to Freud*. Cambridge, Mass: Harvard University Press.

Latour, Bruno. 1983. “Give Me a Laboratory and I Will Raise the World.” Pp. 141–69 in *Science Observed: Perspectives on the Social Study of Science*, edited by K. D. Knorr-Cetina and M. Mulkay. London: Sage.

Latour, Bruno. 2005. *Reassembling the Social*. New York: Oxford University Press.

Laurence, David. 2014. *Opportunity Costs of the PhD: The Problem of Time to Degree*. Modern Language Association.

Le Bon, Gustave. 1879. “Recherches Anatomiques et Mathematiques Sur Les Lois Des Variations Du Volume Du Cerveau et Sur Leurs Relations Avec l’intelligence.” *Revue d’anthropologie* 2(1):27–104.

Leslie, Mitch. 2019. “The Mismeasure of Hands?” *Science* 364(6444):923–25.

Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland. 2015. “Expectations of Brilliance Underlie Gender Distributions across Academic Disciplines.” *Science* 347(6219):262–65. doi: 10.1126/science.1261375.

Linn, Kristin A., Bilwaj Gaonkar, Jimit Doshi, Christos Davatzikos, and Russell T. Shinohara. 2016. “Multivariate Pattern Analysis and Confounding in Neuroimaging.” *The International Journal of Biostatistics* 12(1):31–44. doi: 10.1515/ijb-2015-0030.

Lippa, Richard A. 2006. “The Gender Reality Hypothesis.” *American Psychologist* 61(6):639–40. doi: 10.1037/0003-066X.61.6.639.

Lockhart, Jeffrey W. 2020. “‘A Large and Long Standing Body’: Historical Authority in the Science of Sex.” Pp. 359–86 in *Far Right Revisionism and the End of History: Alt/Histories*, edited by L. D. Valencia-García. New York: Routledge.

Lockhart, Jeffrey W. 2021. “Paradigms of Sex Research and Women in STEM.” *Gender & Society* 35(3):449–75. doi: 10.1177/08912432211001384.

Lockhart, Jeffrey W., and Gary M. Weiss. 2014. “The Benefits of Personalized Smartphone-Based Activity Recognition Models.” Pp. 614–22 in *Proceedings of the 2014 SIAM International Conference on Data Mining*.

Lorenz, Konrad. 1973. “Konrad Lorenz – Biographical.” *NobelPrize.Org*. Retrieved May 27, 2019 (https://www.nobelprize.org/prizes/medicine/1973/lorenz/biographical/).

Lowie, Robert H., and Leta Stetter Hollingworth. 1916. “Science and Feminism.” *The Scientific Monthly* 3(3):277–84.

Lytton, Hugh, and David M. Romney. 1991. “Parents’ Differential Socialization of Boys and Girls: A Meta-Analysis.” *Psychological Bulletin* 109(2):267–96.

Mahan, Margo. 2017. "The Racial Origins of U.S. Domestic Violence Law." UC Berkeley.

Malthus, Thomas Robert. [1798] 1958. *An Essay On Population*. London: J. M. Dent & Sons.

Martin, Emily. 1991. "The Egg and the Sperm: How Science Has Constructed a Romance Based on Stereotypical Male-Female Roles." *Signs* 16(3):485–485. doi: 10.1086/494680.

Mayrl, Damon, and Nicholas Hoover Wilson. 2020. "What Do Historical Sociologists Do All Day? Analytic Architectures in Historical Sociology." *American Journal of Sociology* 125(5):1345–94. doi: 10.1086/709017.

McCabe, David P., and Alan D. Castel. 2008. "Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning." *Cognition* 107(1):343–52. doi: 10.1016/j.cognition.2007.07.017.

McCabe, Janice. 2005. "What's in a Label? The Relationship between Feminist Self-Identification and 'Feminist' Attitudes among U.S. Women and Men." *Gender & Society* 19(4):480–505. doi: 10.1177/0891243204273498.

McQuillan, Dan. 2018. "Data Science as Machinic Neoplatonism." *Philosophy & Technology* 31(2):253–72. doi: 10.1007/s13347-017-0273-3.

Meadow, Tey. 2013. "Queer Numbers: Social Science as Cultural Heterosexism." in *American Sociological Association*. New York.

Mears, Charlotte. 2020. "The Far Right and Women's History." Pp. 91–104 in *Far-Right Revisionism and the End of History: Alt/Histories*, edited by L. D. Valencia-García. New York : Routledge Taylor & Francis Group, [2020] |: Routledge.

Messer-Davidow, Ellen. 2002. *Disciplining Feminism: From Social Activism to Academic Discourse*. Durham: Duke University Press.

Meyer-Bahlburg, Heino F. L. 2005. "Gender Identity Outcome in Female-Raised 46,XY Persons with Penile Agenesis, Cloacal Exstrophy of the Bladder, or Penile Ablation." *Archives of Sexual Behavior* 34(4):423–38.

Miyagi, Miriam, Eartha May Guthman, and Simón(e) Dow-Kuang Sun. 2021. "Transgender Rights Rely on Inclusive Language." *Science* 374(6575):1568–69.

Moè, Angelica. 2012. "Gender Difference Does Not Mean Genetic Difference: Externalizing Improves Performance in Mental Rotation." *Learning and Individual Differences* 22(1):20–24. doi: 10.1016/j.lindif.2011.11.001.

Morning, Ann. 2007. "'Everyone Knows It's a Social Construct': Contemporary Science and the Nature of Race." *Sociological Focus* 40(4):436–54. doi: 10.1080/00380237.2007.10571319.

Murray, Charles. 2009. "Conclusion: Why It All Matters and What Is To Be Done." Pp. 285–300 in *The science on women and science*, edited by C. H. Sommers. Washington, D.C.: AEI Press.

Musto, Michela. 2019. "Brilliant or Bad: The Gendered Social Construction of Exceptionalism in Early Adolescence." *American Sociological Review* 84(3):369–93. doi: 10.1177/0003122419837567.

Nelson, Laura K. 2019. "To Measure Meaning in Big Data, Don't Give Me a Map, Give Me Transparency and Reproducibility." *Sociological Methodology* 49(1):139–43. doi: 10.1177/0081175019863783.

Nelson, Laura K. 2021. "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics* 101539. doi: 10.1016/j.poetic.2021.101539.

Nieuwenhuis, Mireille, Hugo G. Schnack, Neeltje E. van Haren, Julia Lappin, Craig Morgan, Antje A. Reinders, Diana Gutierrez-Tordesillas, Roberto Roiz-Santiañez, Maristela S. Schaufelberger, Pedro G. Rosa, Marcus V. Zanetti, Geraldo F. Busatto, Benedicto Crespo-Facorro, Patrick D. McGorry, Dennis Velakoulis, Christos Pantelis, Stephen J. Wood, René S. Kahn, Janaina Mourao-Miranda, and Paola Dazzan. 2017. "Multi-Center MRI Prediction Models: Predicting Sex and Illness Course in First Episode Psychosis Patients." *NeuroImage* 145:246–53. doi: 10.1016/j.neuroimage.2016.07.027.

NIH. 2015. "NOT-OD-15-102: Consideration of Sex as a Biological Variable in NIH-Funded Research." *National Institutes of Health*. Retrieved November 29, 2020 (https://grants.nih.gov/grants/guide/notice-files/not-od-15-102.html).

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

Ojala, Markus, and Gemma C. Garriga. 2010. "Permutation Tests for Studying Classifier Performance." *Journal of Machine Learning Research* 11:1833–63.

Oudshoorn, Nelly. 2003. *Beyond the Natural Body: An Archaeology of Sex Hormones*. Routledge.

Oudshoorn, Nelly. 2014. "The Birth of Sex Hormones." Pp. 232–62 in *Women and Gender in Science and Technology*. Vol. II, edited by L. Schiebinger. London: Routledge.

Panofsky, Aaron. 2014. *Misbehaving Science: Controversy and the Development of Behavior Genetics*. Chicago: The University of Chicago Press.

Panofsky, Aaron. 2018. "Rethinking Scientific Authority: Behavior Genetics and Race Controversies." *American Journal of Cultural Sociology* 6(2):322–58. doi: 10.1057/s41290-017-0032-z.

Pape, Madeleine. 2021. "Co-Production, Multiplied: Enactments of Sex as a Biological Variable in US Biomedicine." *Social Studies of Science* 0306312720985939. doi: 10.1177/0306312720985939.

Patil, Vrushali. 2018. "The Heterosexual Matrix as Imperial Effect." *Sociological Theory* 36(1):1–26. doi: 10.1177/0735275118759382.

Pfeffer, Carla A. 2018. "Queer Accounting: Methodological Investments and Disinvestments." Pp. 304–25 in *Other: Please Specify*. Oakland: Univ of California Press.

Philosopher, T. 2019. "I Am Leaving Academic Philosophy Because of Its Transphobia Problem." *Medium*. Retrieved January 25, 2020 (https://medium.com/@transphilosopher33/i-am-leaving-academic-philosophy-because-of-its-transphobia-problem-bc618aa55712).

Phoenix, C. H., R. W. Goy, A. A. Gerall, and W. C. Young. 1959. "Organizing Action of Prenatally Administered Testosterone Propionate on the Tissues Mediating Mating Behavior in the Female Guinea Pig." *Endocrinology* 65:163–96.

Pielke, Roger, and Madeleine Pape. 2019. "Science, Sport, Sex, and the Case of Caster Semenya." *Issues in Science and Technology* XXXVI(1):56–63.

Pinker, Steven. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.

Pleasants, Robert K. 2011. "Men Learning Feminism: Protecting Privileges Through Discourses of Resistance." *Men and Masculinities* 14(2):230–50. doi: 10.1177/1097184X11407048.

Poulain de La Barre, François. 1677. *The Woman as Good as the Man, or, The Equallity of Both Sexes*. Cornhill: T.M. for N. Brooks.

Poulian de La Barre, François. 1673. *De l'egalité Des Deux Sexes*. Paris.

Prescod-Weinstein, Chanda. 2020. "Making Black Women Scientists under White Empiricism: The Racialization of Epistemology in Physics." *Signs: Journal of Women in Culture and Society* 45(2):421–47. doi: 10.1086/704991.

van Putten, Michel J. A. M., Sebastian Olbrich, and Martijn Arns. 2018. "Predicting Sex from Brain Rhythms with Deep Learning." *Scientific Reports* 8(1):3069. doi: 10.1038/s41598-018-21495-7.

Quadagno, Jill S. 1979. "Paradigms in Evolutionary Theory: The Sociobiological Model of Natural Selection." *American Sociological Review* 44(1):100–109.

Reinharz, Shulamit. 1986. "Patriarchal Pontifications." *Society* 23(6):23–29.

Rekdal, Ole Bjørn. 2014. "Academic Urban Legends." *Social Studies of Science* 44(4):638–54.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." Pp. 1135–44 in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco.

Richardson, Sarah S. 2013. *Sex Itself: The Search for Male and Female in the Human Genome*. Chicago ; London: The University of Chicago Press.

Richardson, Sarah S. 2022. "Sex Contextualism." *Philosophy, Theory, and Practice in Biology* 14(0). doi: 10.3998/ptpbio.2096.

Richardson, Sarah S., Meredith Reiches, Heather Shattuck-Heidorn, Michelle Lynne LaBonte, and Theresa Consoli. 2015. "Opinion: Focus on Preclinical Sex Differences Will Not Address Women's and Men's Health Disparities." *Proceedings of the National Academy of Sciences* 112(44):13419–20. doi: 10.1073/pnas.1516958112.

Richardson, Sarah S., Meredith W. Reiches, Joe Bruch, Marion Boulicault, Nicole E. Noll, and Heather Shattuck-Heidorn. 2020. "Is There a Gender-Equality Paradox in Science, Technology, Engineering, and Math (STEM)? Commentary on the Study by Stoet and Geary (2018)." *Psychological Science* 4.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford: Oxford University Press.

Rippon, Gina. 2019. *The Gendered Brain: The New Neuroscience That Shatters the Myth of the Female Brain*. London: The Bodley Head.

Risman, Barbara J. 2001. "Calling the Bluff of Value-Free Science." *American Sociological Review* 66(4):605–11. doi: 10.2307/3088926.

Rollins, Oliver. 2021. *Conviction: The Making and Unmaking of the Violent Brain*. Stanford: Stanford University Press.

Rose, Nikolas. 2013. "The Human Sciences in a Biological Age." *Theory, Culture & Society* 30(1):3–34. doi: 10.1177/0263276412456569.

Rosenblatt, Jonathan D. 2016. "Multivariate Revisit to 'Sex beyond the Genitalia.'" *Proceedings of the National Academy of Sciences* 113(14):E1966–67. doi: 10.1073/pnas.1523961113.

Roskies, Adina L. 2008. "Neuroimaging and Inferential Distance." *Neuroethics* 1(1):19–30. doi: 10.1007/s12152-007-9003-3.

Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23(2):325–41.

Roughgarden, Joan. 2004. *Evolution's Rainbow: Diversity, Gender, and Sexuality in Nature and People*. Berkeley: University of California Press.

Roy, Deboleena. 2012. "Cosmopolitics and the Brain: The Co-Becoming of Practices in Feminism and Neuroscience." Pp. 175–92 in *Neurofeminism*, edited by R. Bluhm, A. J. Jacobson, and H. L. Maibom. London: Palgrave Macmillan UK.

Rudin, Cynthia. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1(5):206–15. doi: 10.1038/s42256-019-0048-x.

Salganik, Matt. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Sanchis-Segura, Carla, Naiara Aguirre, Álvaro Javier Cruz-Gómez, Sonia Félix, and Cristina Forn. 2021. "Beyond 'Sex Prediction': Estimating and Interpreting Multivariate Sex Differences and Similarities in the Brain."

Sanchis-Segura, Carla, Maria Victoria Ibañez-Gual, Naiara Aguirre, Álvaro Javier Cruz-Gómez, and Cristina Forn. 2020. "Effects of Different Intracranial Volume Correction Methods on Univariate Sex Differences in Grey Matter Volume and Multivariate Sex Prediction." *Scientific Reports* 10(1):1–15. doi: 10.1038/s41598-020-69361-9.

Sanz, Veronica. 2017. "No Way Out of the Binary: A Critical History of the Scientific Production of Sex." *Signs* 43(1):1–27.

Savic, Ivanka, Alicia Garcia-Falgueras, and Dick F. Swaab. 2010. "Sexual Differentiation of the Human Brain in Relation to Gender Identity and Sexual Orientation." Pp. 41–62 in *Progress in Brain Research*. Vol. 186, *Sex Differences in the Human Brain, their Underpinnings and Implications*, edited by I. Savic. Elsevier.

Sax, Leonard. 2017. *Why Gender Matters: What Parents and Teachers Need to Know about the Emerging Science of Sex Differences*. Second edition. New York: Harmony Books.

Scheuerman, Morgan Klaus, Madeleine Pape, and Alex Hanna. 2021. "Auto-Essentialization: Gender in Automated Facial Analysis as Extended Colonial Project." *Big Data & Society* 8(2):20539517211053710. doi: 10.1177/20539517211053712.

Schilling, Mark F., Ann E. Watkins, and William Watkins. 2002. "Is Human Height Bimodal?" *American Statistician* 56(3):223–29.

Schilt, Kristen. 2010. *Just One of the Guys?: Transgender Men and the Persistence of Gender Inequality*. University of Chicago Press.

Schilt, Kristen, and Laurel Westbrook. 2009. "Doing Gender, Doing Heteronormativity: 'Gender Normals,' Transgender People, and the Social Maintenance of Heterosexuality." *Gender & Society* 23(4):440–64.

Schulz, Marc-Andre, B. T. Thomas Yeo, Joshua T. Vogelstein, Janaina Mourao-Miranada, Jakob N. Kather, Konrad Kording, Blake Richards, and Danilo Bzdok. 2020. "Different Scaling of Linear Models and Deep Learning in UKBiobank Brain Images versus Machine-Learning Datasets." *Nature Communications* 11(1):4238. doi: 10.1038/s41467-020-18037-z.

Schwemmer, Carsten, Carly Knight, Emily D. Bello-Pardo, Stan Oklobdzija, Martijn Schoonvelde, and Jeffrey W. Lockhart. 2020. "Diagnosing Gender Bias in Image Recognition Systems." *Socius* 6:1–17. doi: 10.1177/2378023120967171.

Seaver, Nick. 2019. "Knowing Algorithms." Pp. 412–22 in *digitalSTS: A Field Guide for Science & Technology Studies*, edited by J. Vertesi and D. Ribes. Princeton: Princeton University Press.

Sepehrband, Farshid, Kirsten M. Lynch, Ryan P. Cabeen, Clio Gonzalez-Zacarias, Lu Zhao, Mike D'Arcy, Carl Kesselman, Megan M. Herting, Ivo D. Dinov, Arthur W. Toga, and Kristi A. Clark. 2018. "Neuroanatomical Morphometric Characterization of Sex Differences in Youth Using Statistical Learning." *NeuroImage* 172:217–27. doi: 10.1016/j.neuroimage.2018.01.065.

Serano, Julia. 2008. "A Matter of Perspective: A Transsexual Woman-Centric Critique of Dreger's 'Scholarly History' of the Bailey Controversy." *Archives of Sexual Behavior* 37(3):491–94.

Seron, Carroll, Susan Silbey, Erin Cech, and Brian Rubineau. 2018. "'I Am Not a Feminist, but. . .': Hegemony of a Meritocratic Ideology and the Limits of Critique Among Women in Engineering." *Work and Occupations* 45(2):131–67. doi: 10.1177/0730888418759774.

Shields, Stephanie A. 1980. "Nineteenth-Century Evolutionary Theory and Male Scientific Bias." Pp. 489–502 in *Sociobiology: Beyond Nature/Nurture?*, edited by G. W. Barlow and J. Silverberg.

Shields, Stephanie A. 2016. "Functionalism, Darwinism, and Advances in the Psychology of Women and Gender: From the 19th Century to the 21st." *Feminism & Psychology* 26(4):397–404.

Soh, Debra W. 2017. "The Futility of Gender-Neutral Parenting." *Los Angeles Times*, January 6.

Stacy, Tyler. 2020. "Transforming the Law: Canada's Bill C-16, Gender, and Post-Truth Politics." in *Far-Right Revisionism and the End of History: Alt/Histories*, edited by L. D. Valencia-García. New York : Routledge Taylor & Francis Group, [2020] |: Routledge.

Storage, Daniel, Zachary Horne, Andrei Cimpian, and Sarah-Jane Leslie. 2016. "The Frequency of 'Brilliant' and 'Genius' in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields." *PLOS ONE* 11(3):e0150194. doi: 10.1371/journal.pone.0150194.

Subramaniam, Banu. 2000. "Snow Brown and the Seven Detergents: A Metanarrative on Science and the Scientific Method." *Women's Studies Quarterly* 28(1/2):296–304.

Summers, Lawrence H. 2005. "Remarks at NBER Conference on Diversifying the Science & Engineering Workforce." *Harvard Office of the President*. Retrieved January 14, 2020 (https://web.archive.org/web/20050301023424/http://www.president.harvard.edu/speeches/2005/nber.html).

Taylor, Catherine J. 2016. "'Relational by Nature'? Men and Women Do Not Differ in Physiological Response to Social Stressors Faced by Token Women." *American Journal of Sociology* 122(1):49–89. doi: 10.1086/686698.

Travis, Cheryl Brown. 2003. *Evolution, Gender, and Rape*. MIT Press.

Trujillo, Anelis K., Emily N. Kessé, Oliver Rollins, Sergio Della Sala, and Roberto Cubelli. 2021. "A Discussion on the Notion of Race in Cognitive Neuroscience Research." *Cortex*. doi: 10.1016/j.cortex.2021.11.007.

Tuana, Nancy. 1983. "Re-Fusing Nature/Nurture." *Women's Studies International Forum* 6(6):621–32. doi: 10.1016/0277-5395(83)90023-7.

Tuana, Nancy. 2004. "Coming to Understand: Orgasm and the Epistemology of Ignorance." *Hypatia* 19(1):194–232.

Umek, Wolfgang, and Barbara Fischer. 2020. “We Should Abandon ‘Race’ as a Biological Category in Biomedical Research.” *Female Pelvic Medicine & Reconstructive Surgery* 26(12):719–20. doi: 10.1097/SPV.0000000000000979.

Unger, Rhoda, and Andrea L. Dottolo. 2016. “Historical Significance of Shields’ 1975 Essay: A Brief Commentary on Four Major Contributions.” *Feminism & Psychology* 26(3):272–81.

Valencia-García, Louie Dean. 2020. *Far-Right Revisionism and the End of History: Alt/Histories*. London: Routledge.

Varoquaux, Gaël. 2018. “Cross-Validation Failure: Small Sample Sizes Lead to Large Error Bars.” *NeuroImage* 180:68–77. doi: 10.1016/j.neuroimage.2017.06.061.

Vines, Gail. 1994. *Raging Hormones: Do They Rule Our Lives?* 1st University of California ed. Berkeley: University of California.

Wachinger, Christian, Polina Golland, William Kremen, Bruce Fischl, and Martin Reuter. 2015. “BrainPrint: A Discriminative Characterization of Brain Morphology.” *NeuroImage* 109:232–48. doi: 10.1016/j.neuroimage.2015.01.032.

Wade, Lisa. 2013. “The New Science of Sex Difference.” *Sociology Compass* 7(4):278–93. doi: https://doi.org/10.1111/soc4.12028.

Waidzunas, Tom. 2012. “Young, Gay, and Suicidal: Dynamic Nominalism and the Process of Defining a Social Problem with Statistics.” *Science, Technology, & Human Values* 37(2):199–225. doi: 10.1177/0162243911402363.

Waidzunas, Tom. 2015. *The Straight Line: How the Fringe Science of Ex-Gay Therapy Reoriented Sexuality*. Minneapolis: University of Minnesota Press.

Wang, Lubin, Hui Shen, Feng Tang, Yufeng Zang, and Dewen Hu. 2012. “Combined Structural and Resting-State Functional MRI Analysis of Sexual Dimorphism in the Young Adult Human Brain: An MVPA Approach.” *NeuroImage* 61(4):931–40. doi: 10.1016/j.neuroimage.2012.03.080.

Wechsler, David. 1958. *The Measurement and Appraisal of Adult Intelligence*. Fourth Edition. Baltimore: Williams & Wilkins.

Weeks, Jeffrey. 1990. “The Value of Identity.” Pp. 88–100 in *Identity: Community, Culture, Difference*. London: Lawrence & Wishart.

West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” *Gender and Society* 1(2):125–51. doi: 10.1177/0891243287001002002.

Westbrook, Laurel, and Aliya Saperstein. 2015. “New Categories Are Not Enough: Rethinking the Measurement of Sex and Gender in Social Surveys.” *Gender & Society* 29(4):534–60. doi: 10.1177/0891243215584758.

Wilson, Edward O. [1975] 2000. *Sociobiology: The New Synthesis*. Harvard University Press.

World Rugby. 2020. “Landmark World Rugby Transgender Workshop Important Step towards Appropriate Rugby-Specific Policy.”

Young, Rebecca M., and Evan Balaban. 2006. “Psychoneuroindoctrinology.” *Nature; London* 443(7112):634–634.

Yuan, Lin, Xue Wei, Hui Shen, Ling-Li Zeng, and Dewen Hu. 2018. “Multi-Center Brain Imaging Classification Using a Novel 3D CNN Approach.” *IEEE Access* 6:49925–34. doi: 10.1109/ACCESS.2018.2868813.

Zhang, Yi, Qiang Luo, Chu-Chung Huang, Chun-Yi Zac Lo, Christelle Langley, Sylvane Desrivières, Erin Burke Quinlan, Tobias Banaschewski, Sabina Millenet, Arun L. W. Bokde, Herta Flor, Hugh Garavan, Penny Gowland, Andreas Heinz, Bernd Ittermann, Jean-Luc Martinot, Eric Artiges, Marie-Laure Paillère-Martinot, Frauke Nees, Dimitri Papadopoulos Orfanos, Luise Poustka, Juliane H. Fröhner, Michael N. Smolka, Henrik Walter, Robert Whelan, Shih-Jen Tsai, Ching-Po Lin, Ed Bullmore, Gunter Schumann, Barbara J. Sahakian, Jianfeng Feng, and for the IMAGEN consortium. 2021. “The Human Brain Is Best Described as Being on a Female/Male Continuum: Evidence from a Neuroimaging Connectivity Study.” *Cerebral Cortex* (bhaa408). doi: 10.1093/cercor/bhaa408.

Zoubi, Obada Al, Masaya Misaki, Aki Tsuchiyagaito, Vadim Zotev, Evan White, Tulsa 1000 Investigators, Martin Paulus, and Jerzy Bodurka. 2020. “Predicting Sex from Resting-State FMRI Across Multiple Independent Acquired Datasets.” *BioRxiv* 2020.08.20.259945. doi: 10.1101/2020.08.20.259945.

Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs.