

Introduction

The Power of Algorithms

This book is about the power of algorithms in the age of neoliberalism and the ways those digital decisions reinforce oppressive social relationships and enact new modes of racial profiling, which I have termed *technological redlining*. By making visible the ways that capital, race, and gender are factors in creating unequal conditions, I am bringing light to various forms of technological redlining that are on the rise. The near-ubiquitous use of algorithmically driven software, both visible and invisible to everyday people, demands a closer inspection of what values are prioritized in such automated decision-making systems. Typically, the practice of redlining has been most often used in real estate and banking circles, creating and deepening inequalities by race, such that, for example, people of color are more likely to pay higher interest rates or premiums just because they are Black or Latino, especially if they live in low-income neighborhoods. On the Internet and in our

everyday uses of technology, discrimination is also embedded in computer code and, increasingly, in artificial intelligence technologies that we are reliant on, by choice or not. I believe that artificial intelligence will become a major human rights issue in the twenty-first century. We are only beginning to understand the long-term consequences of these decision-making tools in both masking and deepening social inequality. This book is just the start of trying to make these consequences visible. There will be many more, by myself and others, who will try to make sense of the consequences of automated decision making through algorithms in society.

Part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decisions are made by human beings. While we often think of terms such as “big data” and “algorithms” as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy, which is well documented in studies of Silicon Valley and other tech corridors.

For example, in the midst of a federal investigation of Google’s alleged persistent wage gap, where women are systematically paid less than men in the company’s workforce, an “antidiversity” manifesto authored by

James Damore went viral in August 2017,¹ supported by many Google employees, arguing that women are psychologically inferior and incapable of being as good at software engineering as men, among other patently false and sexist assertions. As this book was moving into press, many Google executives and employees were actively rebuking the assertions of this engineer, who reportedly works on Google search infrastructure. Legal cases have been filed, boycotts of Google from the political far right in the United States have been invoked, and calls for greater expressed commitments to gender and racial equity at Google and in Silicon Valley writ large are under way. What this antidiversity screed has underscored for me as I write this book is that some of the very people who are developing search algorithms and architecture are willing to promote sexist and racist attitudes openly at work and beyond, while we are supposed to believe that these same employees are developing “neutral” or “objective” decision-making tools. Human beings are developing the digital platforms we use, and as I present evidence of the recklessness and lack of regard that is often shown to women and people of color in some of the output of these systems, it will become increasingly difficult for technology companies to separate their systematic and inequitable employment practices, and the far-right ideological bents of some of

their employees, from the products they make for the public.

My goal in this book is to further an exploration into some of these digital sense-making processes and how they have come to be so fundamental to the classification and organization of information and at what cost. As a result, this book is largely concerned with examining the commercial co-optation of Black identities, experiences, and communities in the largest and most powerful technology companies to date, namely, Google. I closely read a few distinct cases of algorithmic oppression for the depth of their social meaning to raise a public discussion of the broader implications of how privately managed, black-boxed information-sorting tools have become essential to many data-driven decisions. I want us to have broader public conversations about the implications of the artificial intelligentsia for people who are already systematically marginalized and oppressed. I will also provide evidence and argue, ultimately, that large technology monopolies such as Google need to be broken up and regulated, because their consolidated power and cultural influence make competition largely impossible. This monopoly in the information sector is a threat to democracy, as is currently coming to the fore as we make sense of information flows through digital media such as Google and Facebook in the wake

of the 2016 United States presidential election.

I situate my work against the backdrop of a twelve-year professional career in multicultural marketing and advertising, where I was invested in building corporate brands and selling products to African Americans and Latinos (before I became a university professor). Back then, I believed, like many urban marketing professionals, that companies must pay attention to the needs of people of color and demonstrate respect for consumers by offering services to communities of color, just as is done for most everyone else. After all, to be responsive and responsible to marginalized consumers was to create more market opportunity. I spent an equal amount of time doing risk management and public relations to insulate companies from any adverse risk to sales that they might experience from inadvertent or deliberate snubs to consumers of color who might perceive a brand as racist or insensitive. Protecting my former clients from enacting racial and gender insensitivity and helping them bolster their brands by creating deep emotional and psychological attachments to their products among communities of color was my professional concern for many years, which made an experience I had in fall 2010 deeply impactful. In just a few minutes while searching on the web, I experienced the perfect storm of insult and injury that I could not

turn away from. While Googling things on the Internet that might be interesting to my stepdaughter and nieces, I was overtaken by the results. My search on the keywords “black girls” yielded HotBlackPussy.com as the first hit.

Hit indeed.

Since that time, I have spent innumerable hours teaching and researching all the ways in which it could be that Google could completely fail when it came to providing reliable or credible information about women and people of color yet experience seemingly no repercussions whatsoever. Two years after this incident, I collected searches again, only to find similar results, as documented in figure I.1.



▶ [Sugary Black Pussy .com-Black girls in a hardcore action galleries](#)
 sugaryblackpussy.com/
 (black pussy and hairy black pussy,black sex,black booty,black ass,black teen pussy,big black ass,black porn star,hot black girl) ...

Figure I.1. First search result on keywords “black girls,” September 2011.

In 2012, I wrote an article for *Bitch* magazine about how women and feminism are marginalized in search results. By August 2012, Panda (an update to Google’s search algorithm) had been released, and pornography was no longer the first series of results for “black girls”; but other girls and women of color, such as

Latinas and Asians, were still pornified. By August of that year, the algorithm changed, and porn was suppressed in the case of a search on “black girls.” I often wonder what kind of pressures account for the changing of search results over time. It is impossible to know when and what influences proprietary algorithmic design, other than that human beings are designing them and that they are not up for public discussion, except as we engage in critique and protest.

This book was born to highlight cases of such algorithmically driven data failures that are specific to people of color and women and to underscore the structural ways that racism and sexism are fundamental to what I have coined *algorithmic oppression*. I am writing in the spirit of other critical women of color, such as Latoya Peterson, cofounder of the blog *Racialicious*, who has opined that racism is the fundamental application program interface (API) of the Internet. Peterson has argued that anti-Blackness is the foundation on which all racism toward other groups is predicated. Racism is a standard protocol for organizing behavior on the web. As she has said, so perfectly, “The idea of a n*gger API makes me think of a racism API, which is one of our core arguments all along—oppression operates in the same formats, runs the same scripts over and over. It

is tweaked to be context specific, but it’s all the same source code. And the key to its undoing is recognizing how many of us are ensnared in these same basic patterns and modifying our own actions.”² Peterson’s allegation is consistent with what many people feel about the hostility of the web toward people of color, particularly in its anti-Blackness, which any perusal of YouTube comments or other message boards will serve up. On one level, the everyday racism and commentary on the web is an abhorrent thing in itself, which has been detailed by others; but it is entirely different with the corporate platform vis-à-vis an algorithmically crafted web search that offers up racism and sexism as the first results. This process reflects a corporate logic of either willful neglect or a profit imperative that makes money from racism and sexism. This inquiry is the basis of this book.

In the following pages, I discuss how “hot,” “sugary,” or any other kind of “black pussy” can surface as the primary representation of Black girls and women on the first page of a Google search, and I suggest that something other than the best, most credible, or most reliable information output is driving Google. Of course, Google Search is an advertising company, not a reliable information company. At the very least, we must ask when we find these kinds of results, Is this the best information? For whom? We must ask

ourselves who the intended audience is for a variety of things we find, and question the legitimacy of being in a “filter bubble,”³ when we do not want racism and sexism, yet they still find their way to us. The implications of algorithmic decision making of this sort extend to other types of queries in Google and other digital media platforms, and they are the beginning of a much-needed reassessment of information as a public good. We need a full-on reevaluation of the implications of our information resources being governed by corporate-controlled advertising companies. I am adding my voice to a number of scholars such as Helen Nissenbaum and Lucas Introna, Siva Vaidhyanathan, Alex Halavais, Christian Fuchs, Frank Pasquale, Kate Crawford, Tarleton Gillespie, Sarah T. Roberts, Jaron Lanier, and Elad Segev, to name a few, who are raising critiques of Google and other forms of corporate information control (including artificial intelligence) in hopes that more people will consider alternatives.

Over the years, I have concentrated my research on unveiling the many ways that African American people have been contained and constrained in classification systems, from Google’s commercial search engine to library databases. The development of this concentration was born of my research training in library and information science. I think of these issues

through the lenses of critical information studies and critical race and gender studies. As marketing and advertising have directly shaped the ways that marginalized people have come to be represented by digital records such as search results or social network activities, I have studied why it is that digital media platforms are resoundingly characterized as “neutral technologies” in the public domain and often, unfortunately, in academia. Stories of “glitches” found in systems do not suggest that the organizing logics of the web could be broken but, rather, that these are occasional one-off moments when something goes terribly wrong with near-perfect systems. With the exception of the many scholars whom I reference throughout this work and the journalists, bloggers, and whistleblowers whom I will be remiss in not naming, very few people are taking notice. We need all the voices to come to the fore and impact public policy on the most unregulated social experiment of our times: the Internet.

These data aberrations have come to light in various forms. In 2015, *U.S. News and World Report* reported that a “glitch” in Google’s algorithm led to a number of problems through auto-tagging and facial-recognition software that was apparently intended to help people search through images more successfully. The first problem for Google was that its photo application had

automatically tagged African Americans as “apes” and “animals.”⁴ The second major issue reported by the *Post* was that Google Maps searches on the word “N*gger”⁵ led to a map of the White House during Obama’s presidency, a story that went viral on the Internet after the social media personality Deray McKesson tweeted it.

These incidents were consistent with the reports of Photoshopped images of a monkey’s face on the image of First Lady Michelle Obama that were circulating through Google Images search in 2009. In 2015, you could still find digital traces of the Google autosuggestions that associated Michelle Obama with apes. Protests from the White House led to Google forcing the image down the image stack, from the first page, so that it was not as visible.⁶ In each case, Google’s position is that it is not responsible for its algorithm and that problems with the results would be quickly resolved. In the *Washington Post* article about “N*gger House,” the response was consistent with other apologies by the company: “Some inappropriate results are surfacing in Google Maps that should not be, and we apologize for any offense this may have caused,’ a Google spokesperson told U.S. News in an email late Tuesday. ‘Our teams are working to fix this issue quickly.’”⁷

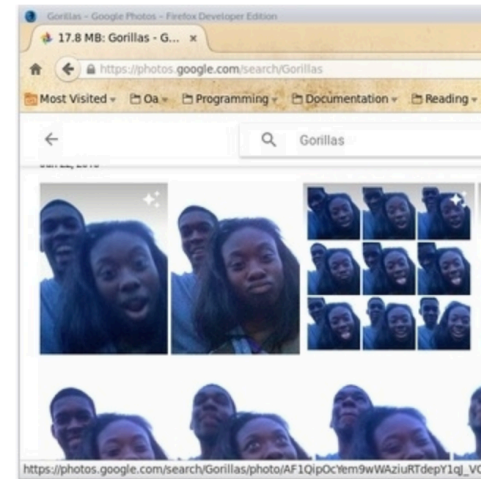


Figure I.2. Google Images results for the keyword “gorillas,” April 7, 2016.

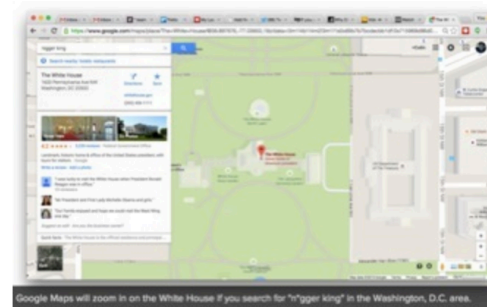


Figure I.3. Google Maps search on “N*gga House” leads to the White House, April 7, 2016.

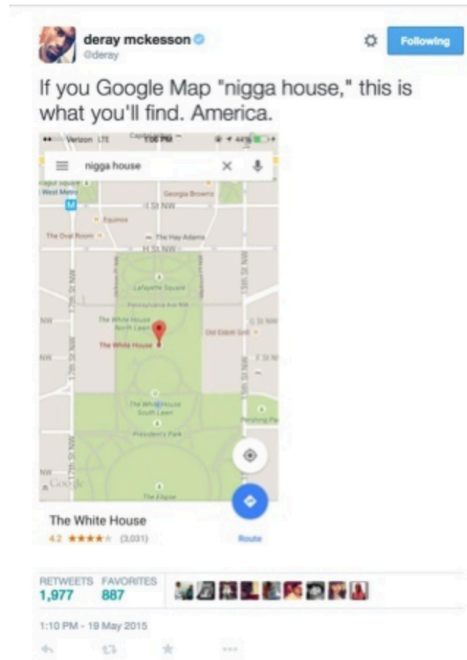


Figure I.4. Tweet by Deray McKesson about Google Maps search and the White House, 2015.



Figure I.5. Standard Google’s “related” searches associates “Michelle Obama” with the term “ape.”

These human and machine errors are not without consequence, and there are several cases that demonstrate how racism and sexism are part of the architecture and language of technology, an issue that needs attention and remediation. In many ways, these cases that I present are specific to the lives and experiences of Black women and girls, people largely understudied by scholars, who remain ever precarious, despite our living in the age of Oprah and Beyoncé in *Shondaland*. The implications of such marginalization are profound. The insights about sexist or racist biases that I convey here are important because information organizations, from libraries to schools and universities to governmental agencies, are increasingly reliant on or being displaced by a variety of web-based “tools” as if there are no political, social, or economic consequences of doing so. We need to imagine new possibilities in the area of information access and knowledge generation, particularly as headlines about “racist algorithms” continue to surface in the media with limited discussion and analysis beyond the superficial.

Inevitably, a book written about algorithms or Google in the twenty-first century is out of date immediately upon printing. Technology is changing rapidly, as are technology company configurations via

mergers, acquisitions, and dissolutions. Scholars working in the fields of information, communication, and technology struggle to write about specific moments in time, in an effort to crystallize a process or a phenomenon that may shift or morph into something else soon thereafter. As a scholar of information and power, I am most interested in communicating a series of processes that have happened, which provide evidence of a constellation of concerns that the public might take up as meaningful and important, particularly as technology impacts social relations and creates unintended consequences that deserve greater attention. I have been writing this book for several years, and over time, Google's algorithms have admittedly changed, such that a search for "black girls" does not yield nearly as many pornographic results now as it did in 2011. Nonetheless, new instances of racism and sexism keep appearing in news and social media, and so I use a variety of these cases to make the point that algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web. It has direct impact on users and on our lives beyond using Internet applications. While I have spent considerable time researching Google, this book tackles a few cases of other algorithmically driven platforms to illustrate how algorithms are serving up

deleterious information about people, creating and normalizing structural and systemic isolation, or practicing digital redlining, all of which reinforce oppressive social and economic relations.

While organizing this book, I have wanted to emphasize one main point: there is a missing social and human context in some types of algorithmically driven decision making, and this matters for everyone engaging with these types of technologies in everyday life. It is of particular concern for marginalized groups, those who are problematically represented in erroneous, stereotypical, or even pornographic ways in search engines and who have also struggled for nonstereotypical or nonracist and nonsexist depictions in the media and in libraries. There is a deep body of extant research on the harmful effects of stereotyping of women and people of color in the media, and I encourage readers of this book who do not understand why the perpetuation of racist and sexist images in society is problematic to consider a deeper dive into such scholarship.

This book is organized into six chapters. In chapter 1, I explore the important theme of corporate control over public information, and I show several key Google searches. I look to see what kinds of results Google's search engine provides about various concepts, and I offer a cautionary discussion of the implications of

what these results mean in historical and social contexts. I also show what Google Images offers on basic concepts such as “beauty” and various professional identities and why we should care.

In chapter 2, I discuss how Google Search reinforces stereotypes, illustrated by searches on a variety of identities that include “black girls,” “Latinas,” and “Asian girls.” Previously, in my work published in the *Black Scholar*,⁸ I looked at the postmortem Google autosuggest searches following the death of Trayvon Martin, an African American teenager whose murder ignited the #BlackLivesMatter movement on Twitter and brought attention to the hundreds of African American children, women, and men killed by police or extrajudicial law enforcement. To add a fuller discussion to that research, I elucidate the processes involved in Google’s PageRank search protocols, which range from leveraging digital footprints from people to the way advertising and marketing interests influence search results to how beneficial this is to the interests of Google as it profits from racism and sexism, particularly at the height of a media spectacle.

In chapter 3, I examine the importance of noncommercial search engines and information portals, specifically looking at the case of how a mass shooter and avowed White supremacist, Dylann Roof, allegedly used Google Search in the development of his

racial attitudes, attitudes that led to his murder of nine African American AME Church members while they worshiped in their South Carolina church in the summer of 2015. The provision of false information that purports to be credible news, and the devastating consequences that can come from this kind of algorithmically driven information, is an example of why we cannot afford to outsource and privatize uncurated information on the increasingly neoliberal, privatized web. I show how important records are to the public and explore the social importance of both remembering and forgetting, as digital media platforms thrive on never or rarely forgetting. I discuss how information online functions as a type of record, and I argue that much of this information and its harmful effects should be regulated or subject to legal protections. Furthermore, at a time when “right to be forgotten” legislation is gaining steam in the European Union, efforts to regulate the ways that technology companies hold a monopoly on public information about individuals and groups need further attention in the United States. Chapter 3 is about the future of information culture, and it underscores the ways that information is not neutral and how we can reimagine information culture in the service of eradicating social inequality.

Chapter 4 is dedicated to critiquing the field of

information studies and foregrounds how these issues of public information through classification projects on the web, such as commercial search, are old problems that we must solve as a scholarly field of researchers and practitioners. I offer a brief survey of how library classification projects undergird the invention of search engines such as Google and how our field is implicated in the algorithmic process of sorting and classifying information and records. In chapter 5, I discuss the future of knowledge in the public and reference the work of library and information professionals, in particular, as important to the development and cultivation of equitable classification systems, since these are the precursors to commercial search engines. This chapter is essential history for library and information professionals, who are less likely to be trained on the politics of cataloguing and classification bias in their professional training. Chapter 6 explores public policy and why we need regulation in our information environments, particularly as they are increasingly controlled by corporations.

To conclude, I move the discussion beyond Google, to help readers think about the impact of algorithms on how people are represented in other seemingly benign business transactions. I look at the “colorblind” organizing logic of Yelp and how business owners are

revolting due to loss of control over how they are represented and the impact of how the public finds them. Here, I share an interview with Kandis from New York,¹⁰ whose livelihood has been dramatically affected by public-policy changes such as the dismantling of affirmative action on college campuses, which have hurt her local Black-hair-care business in a prestigious college town. Her story brings to light the power that algorithms have on her everyday life and leaves us with more to think about in the ecosystem of algorithmic power. The book closes with a call to recognize the importance of how algorithms are shifting social relations in many ways—more ways than this book can cover—and should be regulated with more impactful public policy in the United States than we currently have. My hope is that this book will directly impact the many kinds of algorithmic decisions that can have devastating consequences for people who are already marginalized by institutional racism and sexism, including the 99% who own so little wealth in the United States that the alarming trend of social inequality is not likely to reverse without our active resistance and intervention. Electoral politics and financial markets are just two of many of these institutional wealth-consolidation projects that are heavily influenced by algorithms and artificial intelligence. We need to cause a shift in what



we take for granted in our everyday use of digital media platforms.

I consider my work a practical project, the goal of which is to eliminate social injustice and change the ways in which people are oppressed with the aid of allegedly neutral technologies. My intention in looking at these cases serves two purposes. First, we need interdisciplinary research and scholarship in information studies and library and information science that intersects with gender and women's studies, Black/African American studies, media studies, and communications to better describe and understand how algorithmically driven platforms are situated in intersectional sociohistorical contexts and embedded within social relations. My hope is that this work will add to the voices of my many colleagues across several fields who are raising questions about the legitimacy and social consequences of algorithms and artificial intelligence. Second, now, more than ever, we need experts in the social sciences and digital humanities to engage in dialogue with activists and organizers, engineers, designers, information technologists, and public-policy makers before blunt artificial-intelligence decision making trumps nuanced human decision making. This means that we must look at how the outsourcing of information practices from the public sector facilitates privatization of what

we previously thought of as the public domain²⁴ and how corporate-controlled governments and companies subvert our ability to intervene in these practices.

We have to ask what is lost, who is harmed, and what should be forgotten with the embrace of artificial intelligence in decision making. It is of no collective social benefit to organize information resources on the web through processes that solidify inequality and marginalization—on that point I am hopeful many people will agree.



1

A Society, Searching

On October 21, 2013, the United Nations launched a campaign directed by the advertising agency Mamac Ogilvy & Mather Dubai using “genuine Google searches” to bring attention to the sexist and discriminatory ways in which women are regarded and denied human rights. Christopher Hunt, art director of the campaign, said, “When we came across these searches, we were shocked by how negative they were and decided we had to do something with them.” Kareem Shuhaibar, a copywriter for the campaign, described on the United Nations website what the campaign was determined to show: “The ads are shocking because they show just how far we still have to go to achieve gender equality. They are a wake up call, and we hope that the message will travel far.”¹ Over the mouths of various women of color were the

autosuggestions that reflected the most popular searches that take place on Google Search. The Google Search autosuggestions featured a range of sexist ideas such as the following:

- • Women cannot: drive, be bishops, be trusted, speak in church
- • Women should not: have rights, vote, work, box
- • Women should: stay at home, be slaves, be in the kitchen, not speak in church
- • Women need to: be put in their places, know their place, be controlled, be disciplined

While the campaign employed Google Search results to make a larger point about the status of public opinion toward women, it also served, perhaps unwittingly, to underscore the incredibly powerful nature of search engine results. The campaign suggests that search is a mirror of users’ beliefs and that society still holds a variety of sexist ideas about women. What I find troubling is that the campaign also reinforces the idea that it is not the search engine that is the problem but, rather, the users of search engines who are. It suggests that what is most popular is simply

what rises to the top of the search pile. While serving as an important and disturbing critique of sexist attitudes, the campaign fails to implicate the algorithms or search engines that drive certain results to the top. This chapter moves the lens onto the search architecture itself in order to shed light on the many factors that keep sexist and racist ideas on the first page.



Ad series for UN Women by Mamac Ogilvy & Mather Dubai

Figure 1.1. Mamac Ogilvy & Mather Dubai advertising campaign for the United Nations.

One limitation of looking at the implications of search is that it is constantly evolving and shifting over time. This chapter captures aspects of commercial search at a particular moment—from 2009 to 2015—but surely by the time readers engage with it, it will be a historical rather than contemporary study. Nevertheless, the goal of such an exploration of why

we get troublesome search results is to help us think about whether it truly makes sense to outsource all of our knowledge needs to commercial search engines, particularly at a time when the public is increasingly reliant on search engines in lieu of libraries, librarians, teachers, researchers, and other knowledge keepers and resources.

What is even more crucial is an exploration of how people living as minority groups under the influence of a majority culture, such as people of color and sexual minorities in the United States, are often subject to the whims of the majority and other commercial influences such as advertising when trying to affect the kinds of results that search engines offer about them and their identities. If the majority rules in search engine results, then how might those who are in the minority ever be able to influence or control the way they are represented in a search engine? The same might be true of how men's desires and usage of search is able to influence the values that surround women's identities in search engines, as the Ogilvy campaign might suggest. For these reasons, a deeper exploration into the historical and social conditions that give rise to problematic search results is in order, since rarely are they questioned and most Internet users have no idea how these ideas come to dominate search results on the first page of results in the first place.

Google Search: Racism and Sexism at the Forefront

My first encounter with racism in search came to me through an experience that pushed me, as a researcher, to explore the mechanisms—both technological and social—that could render the pornification of Black women a top search result, naturalizing Black women as sexual objects so effortlessly. This encounter was in 2009 when I was talking to a friend, André Brock at the University of Michigan, who causally mentioned one day, “You should see what happens when you Google ‘black girls.’” I did and was stunned. I assumed it to be an aberration that could potentially shift over time. I kept thinking about it. The second time came one spring morning in 2011, when I searched for activities to entertain my preteen stepdaughter and her cousins of similar age, all of whom had made a weekend visit to my home, ready for a day of hanging out that would inevitably include time on our laptops. In order to break them away from mindless TV watching and cellphone gazing, I wanted to engage them in conversations about what was important to them and

on their mind, from their perspective as young women growing up in downstate Illinois, a predominantly conservative part of Middle America. I felt that there had to be some great resources for young people of color their age, if only I could locate them. I quickly turned to the computer I used for my research (I was pursuing doctoral studies at the time), but I did not let the group of girls gather around me just yet. I opened up Google to enter in search terms that would reflect their interests, demographics, and information needs, but I liked to prescreen and anticipate what could be found on the web, in order to prepare for what might be in store. What came back from that simple, seemingly innocuous search was again nothing short of shocking: with the girls just a few feet away giggling and snorting at their own jokes, I again retrieved a Google Search results page filled with porn when I looked for “black girls.” By then, I thought that my own search history and engagement with a lot of Black feminist texts, videos, and books on my laptop would have shifted the kinds of results I would get. It had not. In intending to help the girls search for information about themselves, I had almost inadvertently exposed them to one of the most graphic and overt illustrations of what the advertisers already thought about them: Black girls were still the fodder of porn sites, dehumanizing them as commodities, as



products and as objects of sexual gratification. I closed the laptop and redirected our attention to fun things we might do, such as see a movie down the street. This best information, as listed by rank in the search results, was certainly not the best information for me or for the children I love. For whom, then, was this the best information, and who decides? What were the profit and other motives driving this information to the top of the results? How had the notion of neutrality in information ranking and retrieval gone so sideways as to be perhaps one of the worst examples of racist and sexist classification of Black women in the digital age yet remain so unexamined and without public critique? That moment, I began in earnest a series of research inquiries that are central to this book.

Of course, upon reflection, I realized that I had been using the web and search tools long before the encounters I experienced just out of view of my young family members. It was just as troubling to realize that I had undoubtedly been confronted with the same type of results before but had learned, or been trained, to somehow become inured to it, to take it as a given that any search I might perform using keywords connected to my physical self and identity could return pornographic and otherwise disturbing results. Why was this the bargain into which I had tacitly entered with digital information tools? And who among us did

not have to bargain in this way? As a Black woman growing up in the late twentieth century, I also knew that the presentation of Black women and girls that I discovered in my search results was not a new development of the digital age. I could see the connection between search results and tropes of African Americans that are as old and endemic to the United States as the history of the country itself. My background as a student and scholar of Black studies and Black history, combined with my doctoral studies in the political economy of digital information, aligned with my righteous indignation for Black girls everywhere. I searched on.



ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM

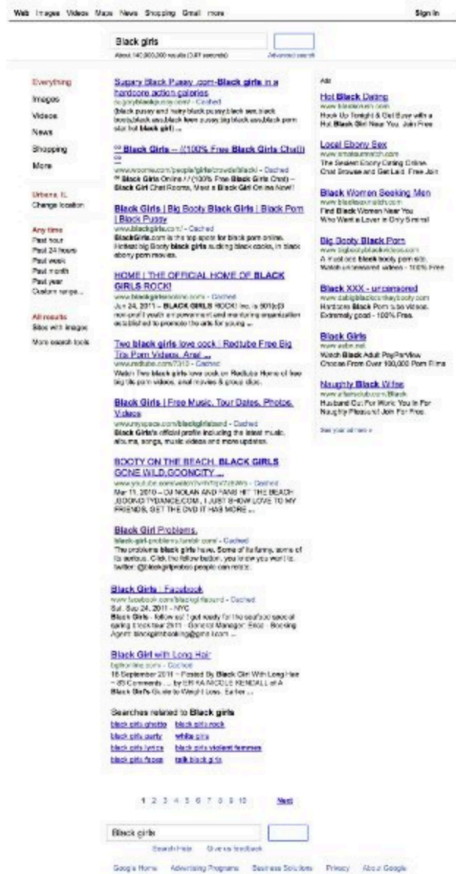


Figure 1.2. First page of search results on keywords “black girls,” September 18, 2011.

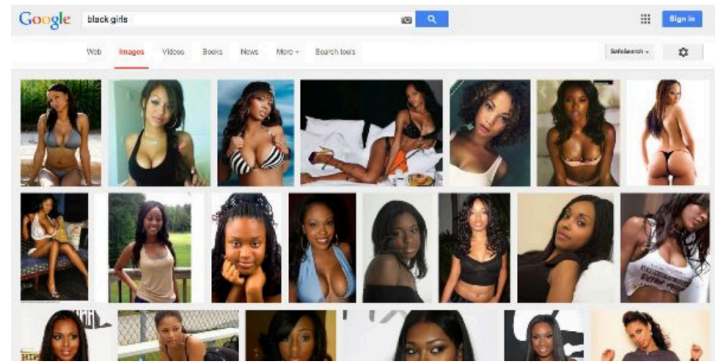


Figure 1.3. First page of image search results on keywords “black girls,” April 3, 2014.

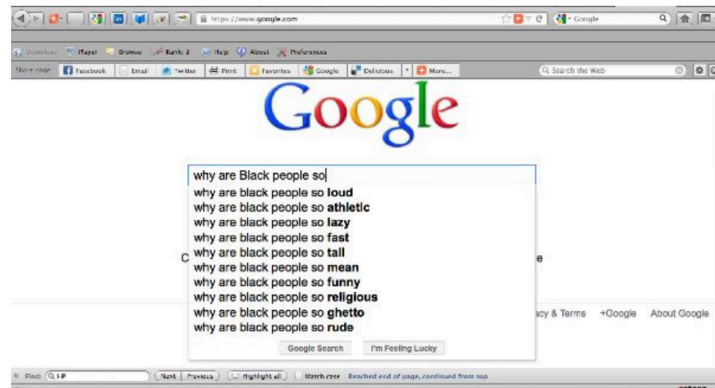


Figure 1.4. Google autosuggest results when searching the phrase “why are black people so,” January 25, 2013.

ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM



Figure 1.5. Google autosuggest results when searching the phrase “why are black women so,” January 25, 2013.

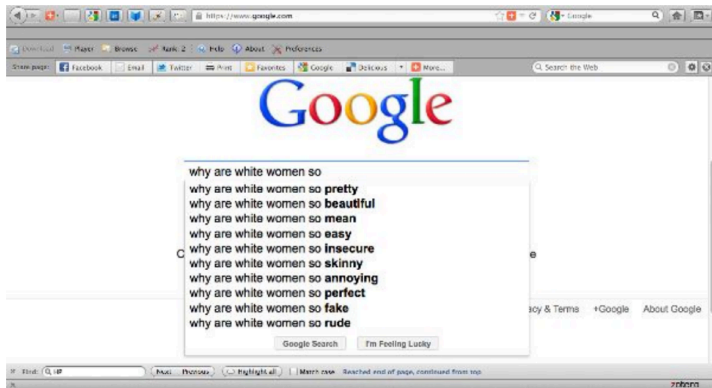


Figure 1.6. Google autosuggest results when searching the phrase “why are white women so,” January 25, 2013.

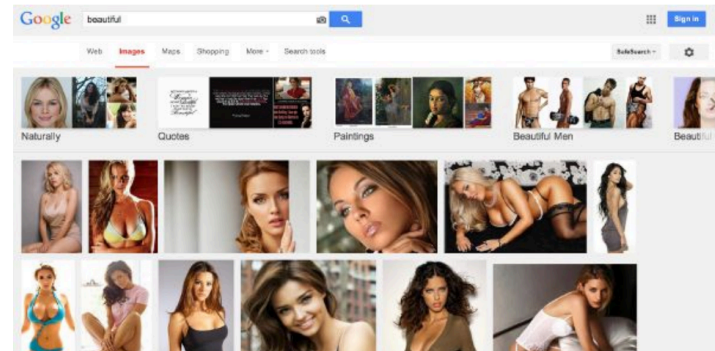


Figure 1.7. Google Images results when searching the concept “beautiful” (did not include the word “women”), December 4, 2014.



Figure 1.8. Google Images results when searching the concept “ugly” (did not include the word “women”), January 5, 2013.

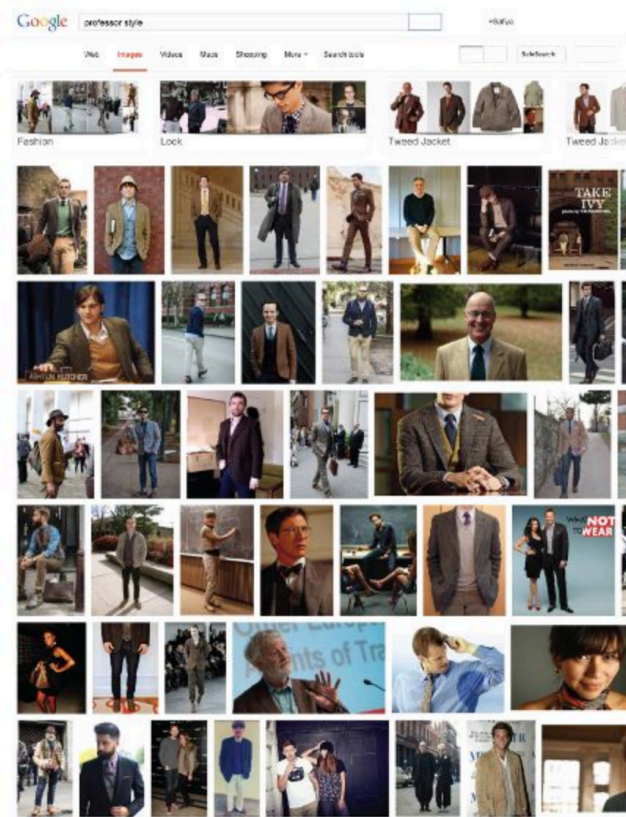


Figure 1.9. Google Images results when searching the phrase “professor style” while logged in as myself, September 15, 2015.

What each of these searches represents are Google’s

algorithmic conceptualizations of a variety of people and ideas. Whether looking for autosuggestions or answers to various questions or looking for notions about what is beautiful or what a professor may look like (which does not account for people who look like me who are part of the professoriate—so much for “personalization”), Google’s dominant narratives reflect the kinds of hegemonic frameworks and notions that are often resisted by women and people of color. Interrogating what advertising companies serve up as credible information must happen, rather than have a public instantly gratified with stereotypes in three-hundredths of a second or less.

In reality, information monopolies such as Google have the ability to prioritize web search results on the basis of a variety of topics, such as promoting their own business interests over those of competitors or smaller companies that are less profitable advertising clients than larger multinational corporations are.² In this case, the clicks of users, coupled with the commercial processes that allow paid advertising to be prioritized in search results, mean that representations of women are ranked on a search engine page in ways that underscore women’s historical and contemporary lack of status in society—a direct mapping of old media traditions into new media architecture. Problematic representations and biases in classifications are not

new. Critical library and information science scholars have well documented the ways in which some groups are more vulnerable than others to misrepresentation and misclassification.³ They have conducted extensive and important critiques of library cataloging systems and information organization patterns that demonstrate how women, Black people, Asian Americans, Jewish people, or the Roma, as “the other,” have all suffered from the insults of misrepresentation and derision in the Library of Congress Subject Headings (LCSH) or through the Dewey Decimal System. At the same time, other scholars underscore the myriad ways that social values around race and gender are directly reflected in technology design.⁴ Their contributions have made it possible for me to think about the ways that race and gender are embedded in Google’s search engine and to have the courage to raise critiques of one of the most beloved and revered contemporary brands.

Search happens in a highly commercial environment, and a variety of processes shape what can be found; these results are then normalized as believable and often presented as factual. The associate professor of sociology at Arizona State University and former president of the Association of Internet Researchers Alex Halavais points to the way that heavily used technological artifacts such as the

search engine have become such a normative part of our experience with digital technology and computers that they socialize us into believing that these artifacts must therefore also provide access to credible, accurate information that is depoliticized and neutral:

Those assumptions are dangerously flawed; . . . unpacking the black box of the search engine is something of interest not only to technologists and marketers, but to anyone who wants to understand how we make sense of a newly networked world. Search engines have come to play a central role in corralling and controlling the ever-growing sea of information that is available to us, and yet they are trusted more readily than they ought to be. They freely provide, it seems, a sorting of the wheat from the chaff, and answer our most profound and most trivial questions. They have become an object of faith.⁵

Unlike the human-labor curation processes of the early Internet that led to the creation of online directories such as Lycos and Yahoo!, in the current Internet environment, information access has been left to the complex algorithms of machines to make selections and prioritize results for users. I agree with

Halavais, and his is an important critique of search engines as a window into our own desires, which can have an impact on the values of society. Search is a symbiotic process that both informs and is informed in part by users. Halavais suggests that every user of a search engine should know how the system works, how information is collected, aggregated, and accessed. To achieve this vision, the public would have to have a high degree of computer programming literacy to engage deeply in the design and output of search.

Alternatively, I draw an analogy that one need not know the mechanism of radio transmission or television spectrum or how to build a cathode ray tube in order to critique racist or sexist depictions in song lyrics played on the radio or shown in a film or television show. Without a doubt, the public is unaware and must have significantly more algorithmic literacy. Since all of the platforms I interrogate in this book are proprietary, even if we had algorithmic literacy, we still could not intervene in these private, corporate platforms.

To be specific, knowledge of the technical aspects of search and retrieval, in terms of critiquing the computer programming code that underlies the systems, is absolutely necessary to have a profound impact on these systems. Interventions such as Black

Girls Code, an organization focused on teaching young, African American girls to program, is the kind of intervention we see building in response to the ways Black women have been locked out of Silicon Valley venture capital and broader participation. Simultaneously, it is important for the public, particularly people who are marginalized—such as women and girls and people of color—to be critical of the results that purport to represent them in the first ten to twenty results in a commercial search engine. They do not have the economic, political, and social capital to withstand the consequences of misrepresentation. If one holds a lot of power, one can withstand or buffer misrepresentation at a group level and often at the individual level. Marginalized and oppressed people are linked to the status of their group and are less likely to be afforded individual status and insulation from the experiences of the groups with which they are identified. The political nature of search demonstrates how algorithms are a fundamental invention of computer scientists who are human beings—and code is a language full of meaning and applied in varying ways to different types of information. Certainly, women and people of color could benefit tremendously from becoming programmers and building alternative search engines that are less disturbing and that reflect and prioritize a

wider range of informational needs and perspectives.

There is an important and growing movement of scholars raising concerns. Helen Nissenbaum, a professor of media, culture, and communication and computer science at New York University, has written with Lucas Introna, a professor of organization, technology, and ethics at the Lancaster University Management School, about how search engines bias information toward the most powerful online. Their work was corroborated by Alejandro Diaz, who wrote his dissertation at Stanford on sociopolitical bias in Google's products. Kate Crawford and Tarleton Gillespie, two researchers at Microsoft Research New England, have written extensively about algorithmic bias, and Crawford recently coorganized a summit with the White House and New York University for academics, industry, and activists concerned with the social impact of artificial intelligence in society. At that meeting, I participated in a working group on artificial-intelligence social inequality, where tremendous concern was raised about deep-machine-learning projects and software applications, including concern about furthering social injustice and structural racism. In attendance was the journalist Julia Angwin, one of the investigators of the breaking story about courtroom sentencing software Northpointe, used for risk assessment by judges to

determine the alleged future criminality of defendants.⁶ She and her colleagues determined that this type of artificial intelligence miserably mispredicted future criminal activity and led to the overincarceration of Black defendants. Conversely, the reporters found it was much more likely to predict that White criminals would not offend again, despite the data showing that this was not at all accurate. Sitting next to me was Cathy O'Neil, a data scientist and the author of the book *Weapons of Math Destruction*, who has an insider's view of the way that math and big data are directly implicated in the financial and housing crisis of 2008 (which, incidentally, destroyed more African American wealth than any other event in the United States, save for not compensating African Americans for three hundred years of forced enslavement). Her view from Wall Street was telling:

The math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions. Nevertheless, many of these models encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed our lives. Like gods, these mathematical models were opaque,

their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer.²

Our work, each of us, in our respective way, is about interrogating the many ways that data and computing have become so profoundly their own “truth” that even in the face of evidence, the public still struggles to hold tech companies accountable for the products and errors of their ways. These errors increasingly lead to racial and gender profiling, misrepresentation, and even economic redlining.

At the core of my argument is the way in which Google biases search to its own economic interests—for its profitability and to bolster its market dominance at any expense. Many scholars are working to illuminate the ways in which users trade their privacy, personal information, and immaterial labor for “free” tools and services offered by Google (e.g., search engine, Gmail, Google Scholar, YouTube) while the company profits from data mining its users. Recent research on Google by Siva Vaidhyanathan, professor of media studies at the University of

Virginia, who has written one of the most important books on Google to date, demonstrates its dominance over the information landscape and forms the basis of a central theme in this research. Frank Pasquale, a professor of law at the University of Maryland, has also forewarned of the increasing levels of control that algorithms have over the many decisions made about us, from credit to dating options, and how difficult it is to intervene in their discriminatory effects. The political economic critique of Google by Elad Segev, a senior lecturer of media and communication in the Department of Communication at Tel Aviv University, charges that we can no longer ignore the global dominance of Google and the implications of its power in furthering digital inequality, particularly as it serves as a site of fostering global economic divides.

However, what is missing from the extant work on Google is an intersectional power analysis that accounts for the ways in which marginalized people are exponentially harmed by Google. Since I began writing this book, Google’s parent company, Alphabet, has expanded its power into drone technology,³ military-grade robotics, fiber networks, and behavioral surveillance technologies such as Nest and Google Glass.⁴ These are just several of many entry points to thinking about the implications of artificial intelligence as a human rights issue. We need to be

concerned about not only how ideas and people are represented but also the ethics of whether robots and other forms of automated decision making can end a life, as in the case of drones and automated weapons. To whom do we appeal? What bodies govern artificial intelligence, and where does the public raise issues or lodge complaints with national and international courts? These questions have yet to be fully answered.

In the midst of Google's expansion, Google Search is one of the most underexamined areas of consumer protection policy,¹⁰ and regulation has been far less successful in the United States than in the European Union. A key aspect of generating policy that protects the public is the accumulation of research about the impact of what an unregulated commercial information space does to vulnerable populations. I do this by taking a deep look at a snapshot of the web, at a specific moment in time, and interpreting the results against the history of race and gender in the U.S. This is only one of many angles that could be taken up, but I find it to be one of the most compelling ways to show how data is biased and perpetuates racism and sexism. The problems of big data go deeper than misrepresentation, for sure. They include decision-making protocols that favor corporate elites and the powerful, and they are implicated in global economic and social inequality. Deep machine learning, which is

using algorithms to replicate human thinking, is predicated on specific values from specific kinds of people—namely, the most powerful institutions in society and those who control them. Diana Ascher,¹¹ in her dissertation on yellow journalism and cultural time orientation in the Department of Information Studies at UCLA, found there was a stark difference between headlines generated by social media managers from the *LA Times* and those provided by automated, algorithmically driven software, which generated severe backlash on Twitter. In this case, Ascher found that automated tweets in news media were more likely to be racist and misrepresentative, as in the case of police shooting victim Keith Lamont Scott of Charlotte, North Carolina, whose murder triggered nationwide protests of police brutality and excessive force.

There are many such examples. In the ensuing chapters, I continue to probe the results that are generated by Google on a variety of keyword combinations relating to racial and gender identity as a way of engaging a commonsense understanding of how power works, with the goal of changing these processes of control. By seeing and discussing these intersectional power relations, we have a significant opportunity to transform the consciousness embedded in artificial intelligence, since it is in fact, in part, a

product of our own collective creation.



Figure 1.10. Automated headline generated by software and tweeted about Keith Lamont Scott, killed by police in North Carolina on September 20, 2016, as reported by the *Los Angeles Times*.

Theorizing Search: A Black Feminist Project

The impetus for my work comes from theorizing Internet search results from a Black feminist perspective; that is, I ask questions about the structure and results of web searches from the standpoint of a Black woman—a standpoint that drives me to ask different questions than have been previously posed about how Google Search works. This study builds on previous research that looks at the ways in which racialization is a salient factor in various engagements with digital technology represented in video games,¹² websites,¹³ virtual worlds,¹⁴ and digital media platforms.¹⁵ A Black feminist perspective offers an opportunity to ask questions about the quality and content of racial hierarchies and stereotyping that appear in results from commercial search engines such as Google's; it contextualizes them by decentering the dominant lenses through which results about Black women and girls are interpreted. By doing this, I am purposefully theorizing from a feminist perspective, while addressing often-overlooked aspects of race in feminist theories of technology. The professor emeritus of science and technology at UCLA Sandra Harding suggests that there is value in identifying a feminist method and epistemology:

Feminist challenges reveal that the questions that are asked—and, even more significantly,

those that are not asked—are at least as determinative of the adequacy of our total picture as are any answers that we can discover. Defining what is in need of scientific explanation only from the perspective of bourgeois, white men’s experiences leads to partial and even perverse understandings of social life. One distinctive feature of feminist research is that it generates problematics from the perspective of women’s experiences.¹⁶

Rather than assert that problematic or racist results are impossible to correct, in the ways that the Google disclaimer suggests,¹⁷ I believe a feminist lens, coupled with racial awareness about the intersectional aspects of identity, offers new ground and interpretations for understanding the implications of such problematic positions about the benign instrumentality of technologies. Black feminist ways of knowing, for example, can look at searches on terms such as “black girls” and bring into the foreground evidence about the historical tendencies to misrepresent Black women in the media. Of course, these misrepresentations and the use of big data to maintain and exacerbate social relationships serve a powerful role in maintaining racial and gender subjugation. It is the persistent

normalization of Black people as aberrant and undeserving of human rights and dignity under the banners of public safety, technological innovation, and the emerging creative economy that I am directly challenging by showing the egregious ways that dehumanization is rendered a legitimate free-market technology project.

I am building on the work of previous scholars of commercial search engines such as Google but am asking new questions that are informed by a Black feminist lens concerned with social justice for people who are systemically oppressed. I keep my eye on complicating the notion that information assumed to be “fact” (by virtue of its legitimation at the top of the information pile) exists because racism and sexism are profitable under our system of racialized capitalism. The ranking hierarchy that the public embraces reflects our social values that place a premium on being number one, and search-result rankings live in this de facto system of authority. Where other scholars have problematized Google Search in terms of its lack of neutrality and prioritization of its own commercial interests, my critiques aim to explicitly address racist and sexist bias in search, fueled by neoliberal technology policy over the past thirty years.

Black Feminism as Theoretical and Methodological Approach

The commodified online status of Black women's and girls' bodies deserves scholarly attention because, in this case, their bodies are defined by a technological system that does not take into account the broader social, political, and historical significance of racist and sexist representations. The very presence of Black women and girls in search results is misunderstood and clouded by dominant narratives of the authenticity and lack of bias of search engines. In essence, the social context or meaning of derogatory or problematic Black women's representations in Google's ranking is normalized by virtue of their placement, making it easier for some people to believe that what exists on the page is strictly the result of the fact that more people are looking for Black women in pornography than anything else. This is because the public believes that what rises to the top in search is either the most popular or the most credible or both.

Yet this does not explain why the word "porn" does not have to be included in keyword searches on "black girls" and other girls and women of color to bring it to the surface as the primary data point about girls and

women. The political and social meaning of such output is stripped away when Black girls are explicitly sexualized in search rankings without any explanation, particularly without the addition of the words "porn" or "sex" to the keywords. This phenomenon, I argue, is replicated from offline social relations and deeply embedded in the materiality of technological output; in other words, traditional misrepresentations in old media are made real once again online and situated in an authoritative mechanism that is trusted by the public: Google. The study of Google searches as an Internet artifact is telling. Black feminist scholars have already articulated the harm of such media misrepresentations:⁴⁸ gender, class, power, sexuality, and other socially constructed categories interact with one another in a matrix of social relations that create conditions of inequality or oppression.

Black feminist thought offers a useful and antiessentializing lens for understanding how both race and gender are socially constructed and mutually constituted through historical, social, political, and economic processes,⁴⁹ creating interesting research questions and new analytical possibilities. As a theoretical approach, it challenges the dominant research on race and gender, which tends to universalize problems assigned to race or Blackness as "male" (or the problems of men) and organizes gender

as primarily conceived through the lenses and experiences of White women, leaving Black women in a precarious and understudied position. Popular culture provides countless examples of Black female appropriation and exploitation of negative stereotypes either to assert control over the representation or at least to reap the benefits of it. The Black feminist scholar bell hooks has written extensively on the ways that neoliberal capitalism is explicitly implicated in misrepresentations and hypersexualization of Black women. hooks's work is a mandate for Black women interested in theorizing in the new media landscape, and I use it as both inspiration and a call to action for other Black women interested in engaging in critical information studies. In total, this research is informed by a host of scholars who have helped me make sense of the ways that technology ecosystems—from traditional classification systems such as library databases to new media technologies such as commercial search engines—are structuring narratives about Black women and girls. In the cases I present, I demonstrate how commercial search engines such as Google not only mediate but are mediated by a series of profit-driven imperatives that are supported by information and economic policies that underwrite the commodification of women's identities. Ultimately, this book is designed to “make it plain,” as we say in

the Black community, just exactly how it can be that Black women and girls continue to have their image and representations assaulted in the new media environments that are not so unfamiliar or dissimilar to old, traditional media depictions. I intend to meaningfully articulate the ways that commercialization is the source of power that drives the consumption of Black women's and girls' representative identity on the web.

While primarily offering reflection on the effects of search-engine-prioritized content, this research is at the same time intended to bring about a deeper inquiry and a series of strategies that can inform public-policy initiatives focused on connecting Black people to the Internet, in spite of the research that shows that cultural barriers, norms, and power relations alienate Black people from the web.²⁰ After just over a decade of focus on closing the digital divide,²¹ the research questions raised here are meant to provoke a discussion about “what then?” What does it mean to have every Black woman, girl, man, and boy in the United States connected to the web if the majority of them are using a search engine such as Google to access content—whether about themselves or other things—only to find results like those with which I began this introduction? The race to digitize cultural heritage and knowledge is important, but it is

often mediated by a search engine for the user who does not know precisely how to find it, much the way a library patron is reliant on deep knowledge and skills of the reference librarian to navigate the vast volumes of information in the library stacks.

The Importance of Google

Google has become a ubiquitous entity that is synonymous for many everyday users with “the Internet” itself. From serving as a browser of the Internet to handling personal email or establishing Wi-Fi networks and broadband projects in municipalities across the United States, Google, unlike traditional telecommunications companies, has unprecedented access to the collection and provision of data across a variety of platforms in a highly unregulated marketplace and policy environment. We must continue to study the implications of engagement with commercial entities such as Google and what makes them so desirable to consumers, as their use is not without consequences of increased surveillance and privacy invasions and participation in hidden labor practices. Each of these enhances the business model of Google’s parent company, Alphabet, and

reinforces its market dominance across a host of vertical and horizontal markets.²² In 2011, the Federal Trade Commission started looking into Google’s near-monopoly status and market dominance and the harm this could cause consumers. By March 16, 2012, Google was trading on NASDAQ at \$625.04 a share, with a market capitalization of just over \$203 billion. At the time of the hearings, Google’s latest income statement, for December 2011, showed gross profit at \$24.7 billion. It had \$43.3 billion cash on hand and just \$6.21 billion in debt. Google held 66.2% of the search engine market industry in 2012. Google Search’s profits have only continued to grow, and its holdings have become so significant that the larger company has renamed itself Alphabet, with Google Search as but one of many holdings. By the final writing of this book in August 2017, Alphabet was trading at \$936.38 on NASDAQ, with a market capitalization of \$649.49 billion.

The public is aware of the role of search in everyday life, and people’s opinions on search are alarming. Recent data from tracking surveys and consumer-behavior trends by the comScore Media Metrix consumer panel conducted by the Pew Internet and American Life Project show that search engines are as important to Internet users as email is. Over sixty million Americans engage in search, and for the most

part, people report that they are satisfied with the results they find in search engines. The 2005 and 2012 Pew reports on “search engine use” reveal that 73% of all Americans have used a search engine, and 59% report using a search engine every day.²³ In 2012, 83% of search engine users used Google. But Google Search prioritizes its own interests, and this is something far less visible to the public. Most people surveyed could not tell the difference between paid advertising and “genuine” results.

If search is so trusted, then why is a study such as this one needed? The exploration beyond that first simple search is the substance of this book. Throughout the discussion of these and other results, I want to emphasize the main point: there is a missing social context in commercial digital media platforms, and it matters, particularly for marginalized groups that are problematically represented in stereotypical or pornographic ways, for those who are bullied, and for those who are consistently targeted. I use only a handful of illustrative searches to underscore the point and to raise awareness—and hopefully intervention—of how important what we find on the web through commercial search engines is to society.

Search Results as Power

Search results reflect the values and norms of the search company’s commercial partners and advertisers and often reflect our lowest and most demeaning beliefs, because these ideas circulate so freely and so often that they are normalized and extremely profitable. Search results are more than simply what is popular. The dominant notion of search results as being both “objective” and “popular” makes it seem as if misogynist or racist search results are a simple mirror of the collective. Not only do problematic search results seem “normal,” but they seem completely unavoidable as well, even though these ideas have been thoroughly debunked by scholars. Unfortunately, users of Google give consent to the algorithms’ results through their continued use of the product, which is largely unavoidable as schools, universities, and libraries integrate Google products into our educational experiences.²⁴

Google’s monopoly status,²⁵ coupled with its algorithmic practices of biasing information toward the interests of the neoliberal capital and social elites in the United States, has resulted in a provision of information that purports to be credible but is actually a reflection of advertising interests. Stated another way, it can be argued that Google functions in the

interests of its most influential paid advertisers or through an intersection of popular and commercial interests. Yet Google's users think of it as a public resource, generally free from commercial interest. Further complicating the ability to contextualize Google's results is the power of its social hegemony.²⁶ Google benefits directly and materially from what can be called the "labortainment"²⁷ of users, when users consent to freely give away their labor and personal data for the use of Google and its products, resulting in incredible profit for the company.

There are many cases that could be made to show how overreliance on commercial search by the public, including librarians, information professionals, and knowledge managers—all of whom are susceptible to overuse of or even replacement by search engines—is something that we must pay closer attention to right now. Under the current algorithmic constraints or limitations, commercial search does not provide appropriate social, historical, and contextual meaning to already overracialized and hypersexualized people who materially suffer along multiple axes. In the research presented in this study, the reader will find a more meaningful understanding of the kind of harm that such limitations can cause for users reliant on the web as an artifact of both formal and informal culture.²⁸ In sum, search results play a powerful role in

providing fact and authority to those who see them, and as such, they must be examined carefully. Google has become a central object of study for digital media scholars,²⁹ due to recognition on these scholars' parts of the power and impact wielded by the necessity to begin most engagements with social media via a search process and the near universality with which Google has been adopted and embedded into all aspects of the digital media landscape to respond to that need. This work is addressing a gap in scholarship on how search works and what it biases, public trust in search, the relationship of search to information studies, and the ways in which African Americans, among others, are mediated and commodified in Google.

To start revealing some of the processes involved, it is important to think about how results appear. Although one might believe that a query to a search engine will produce the most relevant and therefore useful information, it is actually predicated on a matrix of ways in which pages are hyperlinked and indexed on the web.³⁰ Rendering web content (pages) findable via search engines is an expressly social, economic, and human project, which several scholars have detailed. These renderings are delivered to users through a set of steps (algorithms) implemented by programming code and then naturalized as "objective." One of the reasons this is seen as a neutral process is

because algorithmic, scientific, and mathematical solutions are evaluated through procedural and mechanistic practices, which in this case includes tracing hyperlinks among pages. This process is defined by Google's founders, Sergey Brin and Larry Page, as "voting," which is the term they use to describe how search results move up or down in a ranked list of websites. For the most part, many of these processes have been automated, or they happen through graphical user interfaces (GUIs) that allow people who are not programmers (i.e., not working at the level of code) to engage in sharing links to and from websites.³¹

Research shows that users typically use very few search terms when seeking information in a search engine and rarely use advanced search queries, as most queries are different from traditional offline information-seeking behavior.³² This front-end behavior of users appears to be simplistic; however, the information retrieval systems are complex, and the formulation of users' queries involves cognitive and emotional processes that are not necessarily reflected in the system design.³³ In essence, while users use the simplest queries they can in a search box because of the way interfaces are designed, this does not always reflect how search terms are mapped against more complex thought patterns and concepts that users

have about a topic. This disjunction between, on the one hand, users' queries and their real questions and, on the other, information retrieval systems makes understanding the complex linkages between the content of the results that appear in a search and their import as expressions of power and social relations of critical importance.

The public generally trusts information found in search engines. Yet much of the content surfaced in a web search in a commercial search engine is linked to paid advertising, which in part helps drive it to the top of the page rank, and searchers are not typically clear about the distinctions between "real" information and advertising. Given that advertising is a fundamental part of commercial search, using content analysis to make sense of what *actually* is served up in search is appropriate and consistent with the articulation of feminist critiques of the images of women in print advertising.³⁴ These scholars have shown the problematic ways that women have been represented—as sex objects, incompetent, dependent on men, or underrepresented in the workforce³⁵—and the content and representation of women and girls in search engines is consistent with the kinds of problematic and biased ideas that live in other advertising channels. Of course, this makes sense, because Google Search is in fact an advertising platform, not intended to solely

serve as a public information resource in the way that, say, a library might. Google creates advertising algorithms, not information algorithms.

To understand search in the context of this book, it is important to look at the description of the development of Google outlined by the former Stanford computer science graduate students and cofounders of the company, Sergey Brin and Larry Page, in “The Anatomy of a Large-Scale Hypertextual Web Search Engine.” Their paper, written in graduate school, serves as the architectural framework for Google’s PageRank. In addition, it is crucial to also look at the way that citation analysis, the foundational notion behind Brin and Page’s idea, works as a bibliometric project that has been extensively developed by library and information science scholars. Both of these dynamics are often misunderstood because they do not account for the complexities of human intervention involved in vetting of information, nor do they pay attention to the relative weight or importance of certain types of information.³⁶ For example, in the process of citing work in a publication, all citations are given equal weight in the bibliography, although their relative importance to the development of thought may not be equal at all. Additionally, no relative weight is given to whether a reference is validated, rejected, employed, or engaged—

complicating the ability to know what a citation actually *means* in a document. Authors who have become so mainstream as not to be cited, such as not attributing modern discussions of class or power dynamics to Karl Marx or the notion of “the individual” to the scholar of the Italian Renaissance Jacob Burckhardt, mean that these intellectual contributions may undergird the framework of an argument but move through works without being cited any longer. Concepts that may be widely understood and accepted ways of knowing are rarely cited in mainstream scholarship, an important dynamic that Linda Smith, former president of the Association for Information Science and Technology (ASIS&T) and associate dean of the Information School at the University of Illinois at Urbana-Champaign, argues is part of the flawed system of citation analysis that deserves greater attention if bibliometrics are to serve as a legitimating force for valuing knowledge production.



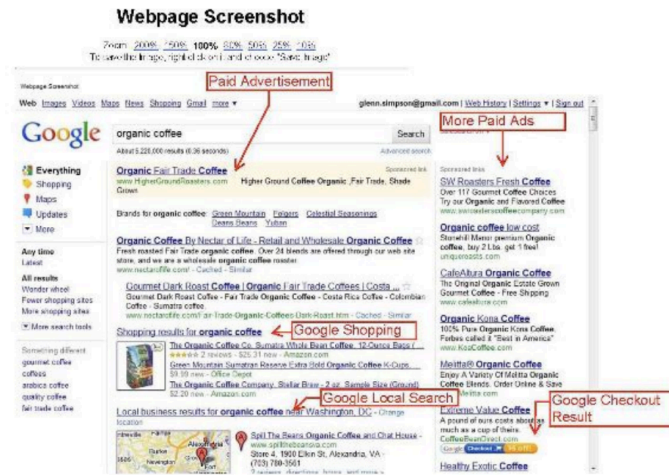


Figure 1.11. Example of Google’s prioritization of its own properties in web search. Source: Inside Google (2010).

Brin and Page saw the value in using works that others cite as a model for thinking about determining what is legitimate on the web, or at least to indicate what is popular based on many people acknowledging particular types of content. In terms of outright co-optation of the citation, vis-à-vis the hyperlink, Brin and Page were aware of some of the challenges I have described. They were clearly aware from the beginning of the potential for “gaming” the system by advertising companies or commercial interests, a legitimated process now known as “search engine optimization,” to

drive ads or sites to the top of a results list for a query, since clicks on web links can be profitable, as are purchases gained by being vetted as “the best” by virtue of placement on the first page of PageRank. This is a process used for web results, not paid advertising, which is often highlighted in yellow (see figure 1.6). Results that appear not to be advertising are in fact influenced by the advertising algorithm. In contrast to scientific or scholarly citations, which once in print are persistent and static, hyperlinking is a dynamic process that can change from moment to moment.³⁷ As a result, the stability of results in Google ranking shifts and is prone to being affected by a number of processes that I will cover, primarily search engine optimization and advertising. This means that results shift over time. The results of what is most hyperlinked using Google’s algorithm today will be different at a later date or from the time that Google’s web-indexing crawlers move through the web until the next cycle.³⁸

Citation importance is a foundational concept for determining scholarly relevance in certain disciplines, and citation analysis has largely been considered a mechanism for determining whether a given article or scholarly work is important to the scholarly community. I want to revisit this concept because it also has implications for thinking about the

legitimation of information, not just citability or popularity. It is also a function of human beings who are engaged in a curation practice, not entirely left to automation. Simply put, if scholars choose to cite a study or document, they have signaled its relevance; thus, human beings (scholars) are involved in making decisions about a document's relevance, although all citations in a bibliography do not share the same level of meaningfulness. Building on this concept of credibility through citation, PageRank is what Brin and Page call the greater likelihood that a document is relevant "if there are many pages that point to it" versus "the probability that the random surfer visits a page."³⁹ In their research, which led to the development of Google Search, Brin and Page discuss the possibility of monopolizing and manipulating keywords through commercialization of the web search process. Their information-retrieval goal was to deliver the most relevant or very best ten or so documents out of the possible number of documents that could be returned from the web. The resulting development of their search architecture is PageRank—a system that is based on "the objective measure of its citation importance that corresponds well with people's subjective idea of importance."⁴⁰

One of the most profound parts of Brin and Page's work is in appendix A, in which they acknowledge the

ways that commercial interests can compromise the quality of search result retrieval. They state, citing Ben Bagdikian, "It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other media, we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers."⁴¹ Brin and Page outline a clear roadmap for how bias would work in advertising-oriented search and the effects this would have, and they directly suggest that it is in the consumer's interest not to have search compromised by advertising and commercialism. To some degree, PageRank was intended to be a measure of relevance based on popularity—including what both web surfers and web designers link to from their sites. As with academic citations, Brin and Page decided that citation analysis could be used as a model for determining whether web links could be ranked according to their importance by measuring how much they were back-linked or hyperlinked to or from. Thus, the model for web indexing pages was born. However, in the case of citation analysis, a scholarly author goes through several stages of vetting and credibility testing, such as the peer-review process, before work can be published



and cited. In the case of the web, such credibility checking is not a factor in determining what will be hyperlinked. This was made explicitly clear in the many news reports covering the 2016 U.S. presidential election, where clickbait and manufactured “news” from all over the world clouded accurate reporting of facts on the presidential candidates.

Another example of the shortcomings of removing this human curation or decision making from the first page of results at the top of PageRank, in addition to the results that I found for “black girls,” can be found in the more public dispute over the results that were returned on searches for the word “Jew,” which included a significant number of anti-Semitic pages. As can be seen by Google’s response to the results of a keyword search for “Jew,” Google takes little responsibility toward the ways that it provides information on racial and gendered identities, which are curated in more meaningful ways in scholarly databases. Siva Vaidhyanathan’s 2011 book *The Googlization of Everything (And Why We Should Worry)* chronicles recent attempts by the Jewish community and Anti-Defamation League to challenge Google’s priority ranking to the first page of anti-Semitic, Holocaust-denial websites. So troublesome were these search results that in 2011, Google issued a statement about its search process, encouraging

people to use “Jews” and “Jewish people” in their searches, rather than the seemingly pejorative term “Jew”—claiming that the company can do nothing about the word’s co-optation by White supremacist groups (see figure 1.12).

Google, according to its own disclaimer, will only remove pages that are considered unlawful, as is the case in France and Germany, where selling or distributing neo-Nazi materials is prohibited. Without such limits on derogatory, racist, sexist, or homophobic materials, Google allows its algorithm—which is, as we can see, laden with what Diaz calls “sociopolitics”—to stand without debate while protesting its inability to remove pages. As recently as June 27, 2012, Google settled a claim by the French antiracism organization the International League Against Racism over Google’s use of ethnic identity —“Jew”—in association with popular searches.⁴² Under French law, racial identity markers cannot be stored in databases, and the auto-complete techniques used in the Google search box link names of people to the word “Jew” on the basis of past user searches. What this recent case points to is another effort to redefine distorted images of people in new media. These cases of distortion, however, continue to accumulate.

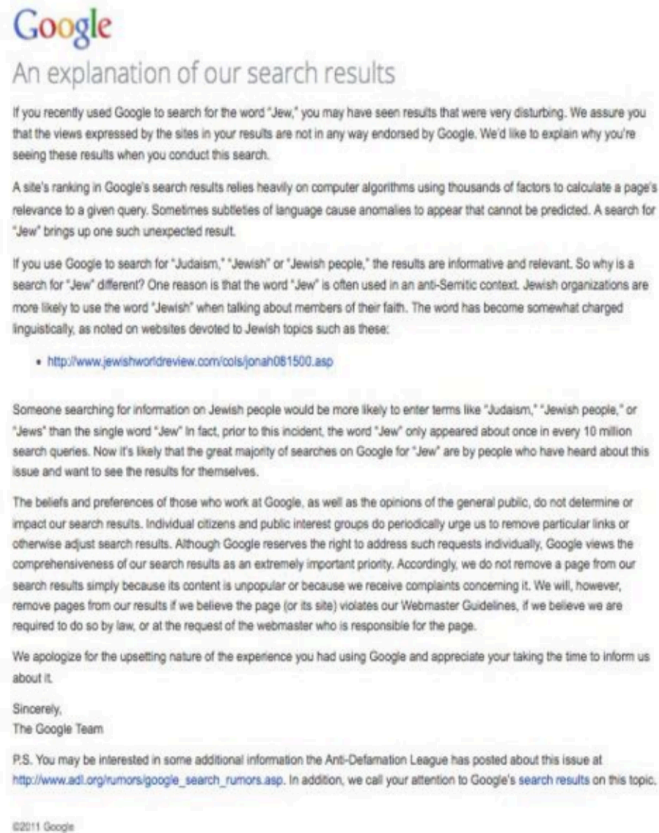


Figure 1.12. Explanation of results by Google. Source: www.google.com/explanation.html (originally available in 2005).

The public's as well as the Jewish community's interest in accurate information about Jewish culture

and the Holocaust should be enough motivation to provoke a national discussion about consumer harm, to which my research shows we can add other cultural and gender-based identities that are misrepresented in search engines. However, Google's assertion that its search results, though problematic, were computer generated (and thus not the company's fault) was apparently a good-enough answer for the Anti-Defamation League (ADL), which declared, "We are extremely pleased that Google has heard our concerns and those of its users about the offensive nature of some search results and the unusually high ranking of peddlers of bigotry and anti-Semitism."⁴³ The ADL does acknowledge on its website its gratitude to Sergey Brin, cofounder of Google and son of Russian Jewish immigrants, for his personal letter to the organization and his mea culpa for the "Jew" search-term debacle. The ADL generously stated in its press release about the incident that Google, as a resource to the public, should be forgiven because "until the technical modifications are implemented, Google has placed text on its site that gives users a clear explanation of how search results are obtained. Google searches are automatically determined using computer algorithms that take into account thousands of factors to calculate a page's relevance."⁴⁴

If there is a technical fix, then what are the

constraints that Google is facing such that eight years later, the issue has yet to be resolved? A search for the word “Jew” in 2012 produces a beige box at the bottom of the results page from Google linking to its lengthy disclaimer about the results—which remain a mix of both anti-Semitic and informative sites (see figure 1.13). That Google places the responsibility for bad results back on the shoulders of information searchers is a problem, since most of the results that the public gets on broad or open-ended racial and gendered searches are out of their control and entirely within the control of Google Search.

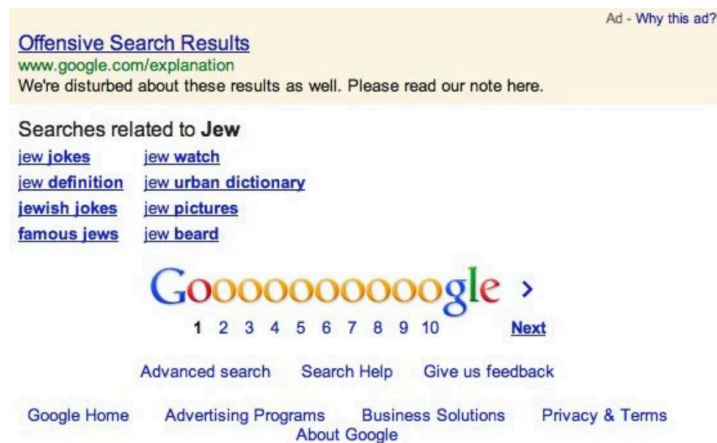


Figure 1.13. Google’s bottom-of-the-page beige box regarding offensive results, which previously took users to “An Explanation of Our Search Results.”

Source: www.google.com/explanation (no longer available).

It is important to note that Google has conceded the fact that anti-Semitism as the primary information result about Jewish people is a problem, despite its disclaimer that tries to put the onus for bad results on the searcher. In Germany and France, for example, it is illegal to sell Nazi memorabilia, and Google has had to put in place filters that ensure online retailers of such are not visible in search results. In 2002, Benjamin Edelman and Jonathan Zittrain at Harvard University’s Berkman Center for Internet and Society concluded that Google was filtering its search results in accordance with local law and precluding neo-Nazi organizations and content from being displayed.⁴⁵ While this indicates that Google can in fact remove objectionable hits, it is equally troubling, because the company provided search results without informing searchers that information was being deleted. That is to say that the results were presented as factual and complete without mention of omission. Yahoo!, another leading U.S. search engine, was forced into a protracted legal battle in France for allowing pro-Nazi memorabilia to be sold through its search engine, in violation of French law. What these cases point to is that search results are deeply contextual and easily

manipulated, rather than objective, consistent, and transparent, and that they can be legitimated only in social, political, and historical context.

The issue of unlawfulness over the harm caused by derogatory results is a question of considerable debate. For example, in the United States, where free speech protections are afforded to all kinds of speech, including hate speech and racist or sexist depictions of people and communities, there is a higher standard of proof required to show harm toward disenfranchised or oppressed people. We need legal protections now more than ever, as automated decision-making systems wield greater power in society.

Gaming the System: Optimizing and Co-opting Results in Search Engines

Google's advertising tool or optimization product is AdWords. AdWords allows anyone to advertise on Google's search pages and is highly customizable. With this tool, an advertiser can set a maximum amount of money that it wants to spend on a daily basis for

advertising. The model for AdWords is that Google will display ads on search pages that it believes are relevant to the kind of search query that is taking place by a user. If a user clicks on an ad, then the advertiser pays. And Google incentivizes advertisers by suggesting that their ads will show up in searches and display, but the advertiser (or Google customer) pays for the ad only when a user (Google consumer) clicks on the advertisement, which is the cost per click (CPC). The advertiser selects a series of "keywords" that it believes closely align with its product or service that it is advertising, and a customer can use a Keyword Estimator tool in order to see how much the keywords they choose to associate with their site might cost. This advertising mechanism is an essential part of how PageRank prioritizes ads on a page, and the association of certain keywords with particular industries, products, and services derives from this process, which works in tandem with PageRank.

In order to make sense of the specific results in keyword searches, it is important to know how Google's PageRank works, what commercial processes are involved in PageRank, how search engine optimization (SEO) companies have been developed to influence the process of moving up results,⁴⁶ and how Google bombing⁴⁷ occurs on occasion. Google bombing is the practice of excessively hyperlinking to a website