Networked Tactics for Gender Representation in the News

by
J. Nathan Matias

Bachelor of Arts in English Literature
Elizabethtown College, 2006

Bachelor of Arts, Honours, Master of Arts, Cantab
University of Cambridge, 2008

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements of the degree of
Master of Science
at the Massachusetts Institute of Technology

June 2013

© 2013 Massachusetts Institute of Technology. All rights reserved.

Signature of Author
Program in Media Arts and Sciences
May 22, 2013

Certified by
Ethan Zuckerman
Director, MIT Center for Civic Media
Principal Research Scientist
MIT Media Lab

Accepted by
Prof Patricia Maes
Associate Academic Head
Program in Media Arts and Sciences
Networked Tactics for Gender Representation in the News

by

J. Nathan Matias

Bachelor of Arts in English Literature
Elizabethtown College, 2006

Bachelor of Arts Honours, Master of Arts, Cantab
University of Cambridge, 2008

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning, on May 22, 2013
in partial fulfillment of the requirements of the degree of
Master of Science
at the Massachusetts Institute of Technology

June 2013

ABSTRACT

This thesis presents research on gender disparities in online news, followed with three open source designs that attempt to address those disparities. Open Gender Tracker is a platform that applies automated gender analysis to electronic content sources. FollowBias is a behavioral experiment on the effectiveness of personal trackers to manage the biases of journalists and curators. Passing On uses data and stories to attract and coordinate participants to expand the visibility of women in Wikipedia. These three designs are offered as inspirations for a paradigm of technologies to measure and change women’s representation in the news.

Thesis supervisor: Ethan Zuckerman
Title: Director, MIT Center for Civic Media,
Principal Research Scientist MIT Media Lab
Networked Tactics for Gender Representation in the News

by
J. Nathan Matias

Advisor
Ethan Zuckerman
Director, MIT Center for Civic Media
Principal Research Scientist
MIT Media Lab

Reader
Tom Steinberg
Director, MySociety

Reader
Kate Crawford
Principal Researcher, Microsoft Research New England
Visiting Professor, MIT Center for Civic Media
ACKNOWLEDGMENTS

Ethan Zuckerman (@ethanz) has been a thoughtful, supportive, and principled advisor, offering context and advice throughout my exploration of civic technology for two years.

Tom Steinberg (@steiny) has offered direct and practical design feedback throughout this project and first suggested that I incorporate randomized controlled trials into my research.

Kate Crawford (@katecrawford) has offered relevant provocations and constructive suggestions throughout the writing process.

Judith Donath (@judithd) doggedly asked important questions that stumped me until I found answers to completely different ones.

Irene Ros (@ireneros) of Bocoup has been the best kind of collaborator, sharing encouragement, inspiration, and rock solid code since the very beginning, especially with Open Gender Tracker, which is primarily her work.

Sarah Szalavitz (@dearsarah) who first proposed FollowBias, has been remarkably prolific and indefatiguable in our collaborative journey to imagine, create, and evaluate what has become a gorgeous and excellent project.

Sophie Diehl has been a talented and capable collaborator on Passing On, creating the beautiful prototypes that inspired it and collaborating closely on this work of meaning and impact.

Adam Hyland (@therealprotonk) of Bocoup defined our approach to citizen media platform analysis and developed the Global Name Gender Data project.

James Home (@jameshome) took the concept of 3D glasses and crafted a beautiful design for FollowBias. I enjoyed the privilege of implementing his elegant design.

Diyang Tang collaborated on the Twitter news quotation project.

David Larochelle (@dlarochelle) collaborated on an afternoon hack that turned into my thesis.

The Knight Foundation funded Open Gender Tracker through a Prototype Fund grant and has supported helpful conversations throughout my work.

Sasha Costanza Chock (@schock) advised my first independent study on this topic and consistently directed me toward principled action grounded in relevant research.

Adrienne Debigare (@adebigare) and Chris Marstall (@marstall) of the Boston Globe Globe Labs supported our integration with the Boston Globe API.

Alberto Ibarguen (@ibarguen) encouraged this project from its very beginning by sending me an interested Tweet that led me to step out my door on this adventure.

Lisa Evans (@objectgroup) collaborated on the study of UK news we produced for the Guardian.

Ami Sedghi (@AmiSedhi) and Simon Rogers (@smfrogers) published our series in the Guardian Datablog.
Anna Powell-Smith (@darkgreener) provided UK name genders for the Guardian series.

Solana Larsen (@solanasaurus) helped us understand Global Voices better.

Emily Bell (@emilybell) offered encouragement, advice, and a chance to brainstorm with students from the Tow Center for Digital Journalism at Columbia Journalism school.

Sam Meier (@sammeier12) helped me navigate the feminist blogosphere.

Matthew of Binders Full of Women answered questions about the Facebook page.

Lynn Cherny (@arnicas) supported data analysis in one of the Guardian posts and inspired the section on women in the book trade.

Brian Keegan (@bkeegan) and SJ Klein (@metasj) helped me navigate Wikipedia’s contribution requirements.

Hanna Wallach (@hannawallach) advised on some of the data analysis.

Katie Orenstein (@katieorenst) and Taryn Yaeger shared ongoing conversations and inspiration on The Op Ed Project and what it means for women to be thought leaders.

Rahul Bhargava (@rahulbot) worked with me on nytcorpus-ruby, a ruby library for processing the historical New York Times content archive.

Charlie De Tar (@cdetar) shared many debugging sessions and feedback about design, consistently encouraging me toward meaningful impact.

Evan Sandhaus (@kansandhaus) and Alexis Lloyd (@alexisloyd) of the New York Times provided encouragement and feedback on our use of the New York Times Corpus.

Bill Thompson (@billt) at the BBC shared the inspiration of his personal approach to gender on speaker panels.

Matt Stempeck (@mstem), Molly Sauter (@mestem), and Kate Darling (@grok) have been inspiring, witty, and encouraging colleagues online and in the office.

All of my other colleagues at the Center for Civic Media: Erhardt Graeff, Catherine D’Ignazio, Lorrie Lejeune, Edward Platt, Leo Burd, Nicole Freedman, Huan Sun, Chris Peterson, Dan Schultz, Denise Cheng, Pablo Rey Mazon, Rogelio Alejandro Lopez, Becky Hurwitz, Andrew Whitacre, Rodrigo Davies, Alexandre Gonclaves, Marco Bani, Willow Brugh, and the Dalek have created a fascinating and supportive community for this work. Except for the Dalek.

danah boyd shared helpful advice whose value will continue long after this thesis.

Jeff Howe took on some of my writing at The Atlantic so I could finish this thesis.

My parents Karin and Jorge Matias continue to encourage and inspire me to do work of principle and empathy.

Hannah Eagleson has shared especially kind encouragement, solidarity, and support.

Thanks also to committee members of the Davies Jackson Scholarship, who maintain an enduring interest in my life, many years after funding my study at the University of Cambridge.
TABLE OF CONTENTS

1. INTRODUCTION ........................................................................................................... 14

2. WOMEN’S REPRESENTATION IN THE MEDIA .......................................................... 16
   2.1 LINKS BETWEEN NEWS COVERAGE AND WOMEN’S POLITICAL PARTICIPATION ... 16
   2.2 EMPLOYMENT DISPARITIES IN NEWSROOMS ...................................................... 17
   2.3 COMMON TACTICS TO ADDRESS WOMEN’S REPRESENTATION IN THE NEWS .... 18
      2.3.1 Regulation ........................................................................................................ 19
      2.3.2 Industry Goals ................................................................................................. 19
      2.3.3 Pressure Campaigns ....................................................................................... 19
   2.4 RELATING WOMEN’S REPRESENTATION TO STRUCTURAL INEQUALITIES ....... 20
      2.4.1 Book Reviews ............................................................................................... 20
      2.4.2 Opinion Writing .............................................................................................. 22
   2.5 ONLINE MEDIA ...................................................................................................... 24
      2.5.1 Blogher ........................................................................................................... 24
      2.5.2 Mommy blogs ............................................................................................... 24
      2.5.3 Feminist blogs .............................................................................................. 25
      2.5.4 Wikipedia ....................................................................................................... 25
      2.5.5 Global Voices and citizen media ..................................................................... 27

3. REPRESENTATION IN NETWORKS .............................................................................. 30
   3.1 MEDIA ACTION FOR REPRESENTATION ............................................................... 30
      3.1.1 Advocacy ....................................................................................................... 30
      3.1.2 Norms .......................................................................................................... 31
      3.1.3 Conversation-starting ................................................................................... 31
      3.1.4 Convening .................................................................................................... 32
      3.1.5 Coordination ............................................................................................... 32
   3.2 NETWORKED POWER ............................................................................................. 33
      3.2.1 Switching power .......................................................................................... 33
      3.2.2 The influence of network actors upon each other .......................................... 34
      3.2.3 Data in networked power ............................................................................ 34
   3.3 REPRESENTATION IN NETWORKS ......................................................................... 35

4. MEASURING GENDER REPRESENTATION IN CONTENT .......................................... 36
   4.1 MANUAL CONTENT ANALYSIS: GLOBAL MEDIA MONITORING PROJECT .......... 36
   4.2 COOPERATIVE ONLINE CONTENT ANALYSIS: PAGEONEX ......................... 38
   4.3 AUTOMATED CONTENT ANALYSIS .................................................................... 39
      4.3.1 Name Gender ................................................................................................ 40
      4.3.2 Subject Gender ............................................................................................ 40
      4.3.3 Language Styles ........................................................................................... 40
   4.4 CHARACTERISTICS OF AUTOMATED CONTENT ANALYSIS .......................... 41
      4.4.1 Variation over time ...................................................................................... 41
      4.4.2 Interpreting automated content analysis in context ..................................... 42
      4.4.3 Filtering and focused exploration .................................................................. 43
      4.4.4 The rate of automated content analysis ...................................................... 45
LIST OF FIGURES

Figure 1: 4thestate chart of gender in 2012 US election coverage ........................................ 17
Figure 2: Opinion writing in New Media, Op Ed Project ...................................................... 23
Figure 3: Opinion writing in legacy media, Op Ed Project .................................................... 23
Figure 4: Global Voices article author gender per year, 2005-2012 ........................................ 28
Figure 5: Global Voices author gender, grouped by posts per author, 2005-2012 ... 28
Figure 6: Functions of news subjects by sex, by occupation: 2010, Global Media Monitoring Project ............................................................... 37
Figure 7: manually-coding news content online: PageOneX ............................................. 39
Figure 8: "Daily Fail" project screenshot: descriptions of women used by all authors, Daily Mail ........................................................................................................... 41
Figure 9: Content author gender and social reach per week, Jul 2011 - June 2012 .. 44
Figure 10: Obituary subject gender per month, New York Times, 1987-2007 ...................... 43
Figure 11: Editorial obituaries, New York Times 1987-2007 .............................................. 43
Figure 12: Paid death notices, New York Times 1987-2007 .............................................. 43
Figure 13: Slopegraph chart of most prolific Guardian bylines, July 2011 - June 2012 .. 44
Figure 14: NYTWrites, by Irene Ros, which illustrates a writer's topic history (Windsheimer) ........................................................................................................... 45
Figure 15: Comparing content gender to social shares UK opinion writing, July 2011 to June 2012 ........................................................................................................... 48
Figure 16: Guardian author gender and social reach per week, Jul 2011 - Jul 2012 .... 49
Figure 17: Global Voices social media sourcing practices, 2005 – Jul 2012 ..................... 53
Figure 18: Ranking of most quoted Twitter sources, Global Voices Egypt ...................... 54
Figure 19: source citation record, Global Voices ............................................................... 55
Figure 20: Dashboard of Twitter accounts quoted in the news ........................................... 56
Figure 21: Nohomophobes.com, with Twitter photos and account names blanked out ... 62
Figure 22: Tweet by Adria Richards ...................................................................................... 63
Figure 23: Open Gender Tracker system architecture .......................................................... 68
Figure 24: Author gender in Global Voices, 2005-2012 ..................................................... 70
Figure 25: Author gender in Global Voices at different levels of participation ............... 71
Figure 26: Author gender in Global Voices across sections .............................................. 71
Figure 27: Open Gender Tracker Interactive API Explorer ................................................. 73
Figure 28: FollowBias user experience .............................................................................. 79
Figure 29: FollowBias system architecture ....................................................................... 80
Figure 30: FollowBias presents results as a pie chart labeled with 3D glasses ............... 80
Figure 31: FollowBias gender corrections interface. Users click or press a circle to make a correction ................................................................................................. 82
Figure 32: Percentage of corrections per study participant where corrections > 0 ...... 84
Figure 33: Personal account FollowBias score .................................................................... 85
Figure 34: Professional account FollowBias score ............................................................ 85
Figure 35: Difference between women % and men % across all participants ............... 87
Figure 36: Change in difference between women % and men % before and after, control group ......................................................................................... 88
Figure 37: Change in difference between women % and men % before and after, treatment group ................................................................. 88
Figure 38: Passing On, by J. Nathan Matias and Sophie Diehl ......................................... 91
Figure 39: how Publication B can include more women in comparison to Publication B, which is smaller, and still be less equal in the balance of who it omits.................. 93
Figure 40: Gender in Memoriam, by Sophie Diehl .......................... 94
Figure 41: Passing On user experience......................................... 95
Figure 42: Passing On system architecture........................................ 97
1. INTRODUCTION

Invisibility is always surprising when you first notice it. When I first saw data on the limited place for women’s voices in the news, I was shocked and incredulous. Doubting the methods used by others, I developed more accurate techniques for measuring large datasets of news content. Gender disparities in the news become even more prominently visible under the lens of those technologies. They also illustrate the role that each of us plays in those disparities every time we click, share, and use our own voices.

As the news industry adapts and evolves online, initiatives towards fairness in the media need to evolve as well. Pressure and awareness campaigns occasionally influence powerful organizations, but the power of those organizations is becoming more diffuse in the network of voices. In this thesis, I present three designs that attempt to support longer-term change. Open Gender Tracker is a platform for monitoring gender in online content. FollowBias tracks personal bias in Twitter behaviour and measures change. Passing On supports constructive collective action towards improving Wikipedia’s inclusion of women. I hope that these open source technologies inspire further tactics to support diverse conversations online.

The first part of this thesis sets the context for my designs by addressing two parallel issues, women’s representation in the media and representation in online networks. Women’s opportunities and aspirations are linked with media cultures that often do not include their voices, across newsrooms, literary publishing, and some online platforms (ch2). Although women’s writing is flourishing in some parts of the Internet, a shift from broadcast media to networks has changed the structure of representation itself, making data on the state of women’s representation incomplete without the use of computational technologies to monitor those networks (ch3). In the case of women’s representation, these technologies can measure content by and about women at large scales and high speeds (ch4). With that content gender data, it is possible to trace women’s representation through the signals that follow content and attention across networks (ch5). Many tactics for change online tend towards awareness and pressure rather than addressing these systemic biases (ch6).
The second part of this thesis presents three working designs that attempt to address systemic gender biases in online news. Open Gender Tracker is a platform that applies automated gender analysis to electronic content sources (ch7). FollowBias is a behavioral experiment on the effectiveness of personal trackers to manage the biases of journalists and curators (ch8). Passing On uses data and stories to attract and coordinate participants to expand the visibility of women in Wikipedia (ch9). Finally, these three technologies are offered as inspiration for a paradigm of technologies to measure and change women’s representation in the news (ch10).
2. WOMEN’S REPRESENTATION IN THE MEDIA

In this section, I explore the connection between women’s political participation and representation in the news. This includes the issues of political awareness and running for office. Women’s employment in newsrooms and coverage of women in the news has been the subject of many years of industry and activist efforts for change. Book reviews and opinion articles are two cases where women’s representation has particularly clear implications outside newsrooms. I also explore cases of women’s participation and representation in online media.

2.1 Links between news coverage and women’s political participation

The representation of women in the news is a critical link in a cycle of political under-representation for women, a cycle which includes voter participation, news coverage, political knowledge, and political candidacy. In a fair society, women would run for office in roughly equal numbers to men, reach parity in representative bodies, receive similar news coverage as male candidates, and receive votes from men and women who are equally informed about politics. The reality is far from this imaginary ideal. Women constitute 51% of US and UK population but hold only 18.1% of the seats in the US congress and less than 25% of all seats in the British Parliament (Center for American Women and Politics). Women are more likely to vote or sign a petition in both countries, but they are much less likely to run for office than men (Electoral Commission; US Census Bureau). According to a study by Jennifer Lawless at American University, women's concerns about media bias and, dealing with the press, and the pressure to run a negative campaign are major deterrents to running for office.

Beyond media treatment of women candidates, there exists a parallel challenge around the political knowledge of women voters. A report by the Shorenstein center summarizes two decades of research on women's political knowledge, showing that women are less likely than men to read political news and also have less political knowledge than men. This difference has been linked to a whole series of structures that reinforce gender inequality: socialisation at an early age, social expectations on the local and community role of women, education inequities, and employment inequities. Yet the differences
between men and women's political knowledge are reduced in regions where women hold or run for public office and receive coverage in the news (Shorenstein, 32-34).

Women's representation in the media is part of overall political awareness, not just for female audiences. Even so, in many cases, women are a minority of who's quoted on topics about women. During the 2012 US election, media monitoring company 4thEstate published reports on how many women were quoted on political topics. They claimed that in six months of newspaper, television, and radio coverage, quotes from men constituted 81% of quotes in stories about abortion and 75% of quotes in stories on birth control. According to 4thestate, women supplied only 31% of quotes in stories about women's rights. Since the news shapes everyone's awareness on issues of common importance, gender disparities in news coverage leave everyone without a full chance to hear from people most affected by these issues.

Figure 1: 4thestate chart of gender in 2012 US election coverage

2.2 Employment disparities in newsrooms

Women's role in the news is also an issue of employment equality. As jobs across the US news industry have declined sharply, women have been more affected than men by that
decline. According to the American Society of Newsroom Editors, around 10,219 women were employed as reporters in American newsrooms in 2001, accounting for nearly 40% of reporters. By 2012, women held 3338 fewer jobs as reporters, constituting around 38% of reporters in newsrooms across the US. Although the entire news industry has experienced very large losses, they have been especially been felt by women, who were already a minority in newsrooms.

Women's employment in the news is correlated with broader coverage of women in newspapers, since research has shown a link between newsroom gender and the gender of sources and news subjects. In studies of a single topic in national news or a single day within regional newspapers, women have been shown to be more likely than men to include women as sources in the news (Armsom, 2004; Freedman, 2005). In a study spanning 1281 newspapers, radio stations, and television stations in 108 countries, the Global Media Monitoring Project has found that women are more likely than men to report stories about women (GMMP 2012, 27). Although these studies have a very limited scope and the findings won't be true for every publication, these studies suggest that media diversity as a whole benefits from more diverse voices in the news when women are part of diverse news teams.

When women are part of newsrooms, they often find themselves writing about entertainment, lifestyle, and general news, according to a study of an entire year's UK news I conducted with Lisa Evans at the Guardian. From July 2011 through June 2012, it was more rare to read a sports, business, or science article by a woman in the Guardian, Telegraph, and Daily Mail. These differences are most often explained and excused by the argument that they reflect structural inequalities in the rest of society. Since women are a minority of professional athletes, there is less sports news about women than men, and it's thus unsurprising that very few sports reporters are women.

2.3 Common tactics to address women’s representation in the news
Attempts to address the representation of women in the mainstream news industry take a variety of forms. Legal challenges attempt to regulate news content and employment.
Industry associations pledge to goals, monitor progress, and support change. Pressure campaigns set out to influence journalists and editors around specific issues. Many of these initiatives are supported by byline counts and content analysis, which is often used as its own form of pressure campaign.

2.3.1 Regulation
Regulators offer one means to change women's representation in the news, if they could be required to enforce standards on the portrayal of women in the media. In the UK, a coalition of the advocacy groups Violence Against Women, Object, Eaves, and Equality Now proposed a complaints commission with investigative ability and the authority to impose sanctions, special protection for children, funding for diversity training, the inclusion of school curriculum about media representation, and professional codes of practice (EqualitNow). Although none of their suggestions were implemented, the official report of the Leveson Inquiry included a suggestion that media regulators should have "the power to take complaints from representative women's groups" (Margolis et al).

2.3.2 Industry Goals
Professional membership societies such as the American Society of News Editors (ASNE) have been another avenue through which newsroom diversity has been attempted. In 1978, ASNE set the goal of reaching demographic parity across the papers where its editors were members, with a special focus on racial diversity. Participating editors support an annual newsroom diversity census, pledge to support diverse recruiting, and work together on progress benchmarks. In 2000, with America's newsrooms still white and male, the ASNE recommitted to reach this goal by 2025.

2.3.3 Pressure Campaigns
Pressure campaigns such as Colorlines's "Drop the I Word" set out to influence journalists and editors by publicising problems in representation and complaining directly to those responsible. Participants in the pledge sign up to be notified when news outlets use the term "illegal immigrant," and coordinate telephone calls to their local media outlets to complain. By reaching out to geographically-diverse participants over
Facebook, Colorlines tries to coordinate a broad media monitoring and pressure campaign. At the time of writing, the campaign had 4,525 Facebook participants.

The success of the above initiatives is evaluated through data collection, and publication of data on diversity is itself sometimes seen as a tactic for change. In March 2013, the organization VIDA, which supports women in the literary arts, complained that in the three years since they started publishing data, very few publications have changed the percentage of women writing or women's books which they review. By publishing data and spreading that data publicly, they hope to raise awareness and thereby influence the behaviour of publishers. Although some publishers have responded positively, VIDA's 2013 press release expressed disappointment at the impact of their approach:

*I fear the attention we’ve already given them has either motivated their editors to disdain the mirrors we’ve held up to further neglect or encouraged them to actively turn those mirrors into funhouse parodies at costs to women writers as yet untallied.*

### 2.4 Relating women’s representation to structural inequalities

Just as women's role in sports writing is connected to the structural inequalities of professional sport, each kind of news interacts with the state of gender in the areas it covers. These inequalities are often put forward as an explanation for the inequalities in newsrooms. Although injustice is never an excuse for more injustice, it's only fair to scrutinise and understand the connection between specific parts of the news and the structures they reflect. Indeed, in the case of book reviews and opinion writing, more focused analysis is a necessary part of planning effective interventions for change.

#### 2.4.1 Book Reviews

When less than 35% of books reviewed by the Times Literary Supplement or the New York Times Book Review were written by women, the social and economic implications are obvious. When a publication chooses to review a book, they channel attention to that book, influence its sales, and wield considerable influence on the career of the writer who's work has been reviewed. Why shouldn't these publications review more women's
writing? On one hand, reviewers have a responsibility to accurately reflect the state of the book trade. Reviewers also have a responsibility to readers, to find the most newsworthy books to review. Reviewers also have an ethical responsibility to wield their influence on the trade responsibly; it's reasonable to use the reviewer's power to support a healthier, more diverse book industry. Yet even this relatively simple equation seems a delicate and complicated balance to make.

Actual data on the state of the book trade shows that this kind of complicated hand wringing may be unnecessary. In the UK, women write the majority of top bestselling fiction, even though women's books are a minority of what gets reviewed. According to data analysis of 23 years of the Nielsen Bookscan ratings by Lynn Cherny, within fiction, women wrote all the bestselling romance novels and men wrote all the crime novels and thrillers. It's understandable to see gender differences in those reviews. Elsewhere, women wrote far more bestselling fantasy and science fiction novels than men. Women and men wrote bestselling literary fiction in nearly equal numbers, and yet men's books are reviewed much more frequently across dozens of literary magazines. In the case of book reviews, ethics of accuracy and newsworthiness tug reviewers to pay more attention to women's writing, not as a social intervention but as a portrayal of the reality of the industry.

Book blogs also exhibit disparities in the gender of whose books they review. Although women write slightly more top bestsellers in science fiction and fantasy, SF/F blogs write about books by women only 42% of the time, according to a study of 25 blogs by LadyBusiness. Just as in journalism, LadyBusiness found that women reviewers reviewed women's books 58% of the time, group blogs only reviewed women's books a third of the time, and blogs by men only reviewed women's writing 25% of the time. Although women online are more likely to review women's writing, even online, book reviews don't match the gender ratios of the book trade.
2.4.2 Opinion Writing

Opinion writing participates in a more direct cycle of power than literary reviews. Op Ed articles are often submitted articles. Editors accept proposals from anyone and choose which submissions to publish. By publishing opinion articles, writers establish themselves as capable and knowledgeable voices on the issues they address. The attention they achieve is a means to organise others around a common interest or a step towards speaking engagements, book contracts, funding, and even government positions. Opinion editors arbitrate this power by choosing whose submissions to publish.

Although 20% of US opinion writing is by women, the percentage of submissions by women is sometimes even less, according to an article in the Columbia Journalism Review by Erika Fry. In 2008, the op-ed editor of the Washington Post reported that ten percent of the Post's opinion article submissions were from women. Fry reports anecdotally that women tend to submit articles based within their expertise, while men are more likely to submit articles based on their "dinner-party" opinions. The small supply of women's submissions puts editors in a difficult position, even when they want to feature more women's writing.

Opinion sections also offer a microcosm of areas in society where women speak and are expected to speak. When Taryn Yaeger of the Op Ed Project counted bylines in four months of top US newspapers in 2011, she found that women wrote a majority of opinion articles on family and gender. Men wrote the majority of articles on everything else, from food and social issues to politics and economics (Yaeger; Fry).
Figure 2: Opinion writing in New Media, Op Ed Project

Figure 3: Opinion writing in legacy media, Op Ed Project
Since opinion sections accept unsolicited submissions, and since women submit a minority of opinion pieces, changes in opinion writing can be supported outside of newsrooms. Instead of using data to call out news organisations, The Op Ed Project uses data on opinion writing to inform its US-wide network of training and mentorship. In an interview, Op Ed Project founder Katie Orienstein explained that the organisation focuses on empowerment beyond just getting articles published. They provide mentorship and training with the aim of supporting women to use their public voice as one step towards achieving their goals.

2.5 Online Media
In principle, the Internet lowers the cost of political speech, fostering new conversations outside mainstream media and opening opportunities for underrepresented people. Perhaps the visibility of women's voices could be changed without changing mainstream media. Online, so called “mommy blogs” and feminist blogs offer alternative spaces for conversations that don't happen in mainstream media. Wikipedia and other volunteer sites don’t have the same kinds of barriers to participation as newsrooms, but are women better represented on these sites than mainstream media?

2.5.1 Blogher
In the US, the largest women's blogging community is BlogHer, a network of around 3,000 women bloggers who syndicate content out to BlogHer and participate in an advertising pool. BlogHer has hosted a series of women's blogger conferences since 2005. According to BlogHer's 2012 Women & Social Media study, women trust blogs more than Facebook, Twitter, or Pinterest. Blogher participants write about topics including careers, entertainment, family, feminism, and food. Since Blogher’s model is centered around common advertising the site is presented to newcomers as a source of information for advice, opinions, and product recommendations.

2.5.2 Mommy blogs
So called mommy bloggers have been recognised as an important online bloc for nearly a decade. In the UK, parenting website Mumsnet is is a major hub of the UK blogosphere.
In the 2009 parliamentary election, candidates reached out directly to Mumsnet (Barnett). In 2011, Mumsnet initially supported Tory proposals for internet content filtering but retracted that support after technically-knowledgeable Mumsnet members objected (Barnett, Williams). Prime Minister David Cameron once responded personally to Mumsnet criticism on cuts to social services ("PM criticised").

2.5.3 Feminist blogs
Feminist blogs like Feministing, Racialicious, Jezebel participate in a bold, pop-culture-infused conversation among women online. In a New York Magazine review, Emily Nussbaum writes that online writing offers power unavailable in the mainstream media:

Freed from the boundaries of print, writers could blur the lines between formal and casual writing; between a call to arms, a confession, and a stand-up routine—and this new looseness of form in turn emboldened readers to join in, to take risks in the safety of the shared spotlight."

Despite vigorous online conversations about parenting and feminism, Internet media may not be broadening women's voices beyond categories where women are expected to write. In the Op Ed project's analysis of new media, sites like the Huffington Post and Salon included more women's writing on topics of gender, food, family, style, and health, but men wrote the majority of everything else (Yaeger). Alternative women's media also struggles to remain sustainable. In January 2013, Vanessa Valenti left Feministing which she co-founded, to start the PR firm ValentiMartin Media. In her farewell post, she argues that "the largest challenge facing online feminist work today [is] that it's completely unsustainable." Like most parts of the content business, feminist blogs can't fund their writers very well, if at all.

2.5.4 Wikipedia
Although some online communities restructure participation in mediamaking, these new cultures and structures don't always exhibit diverse gender representation. For example, Wikipedia is an example of a new kind of economic system enabled by digital networks,
what Yochai Benkler calls “commons-based peer production.” Wikipedians participate in peer production when they voluntarily edit part of an article for the common good or for "self-definition" (Benkler, 1). Since Wikipedians can be anonymous or use pseudonyms and often contribute for free, Wikipedia lacks the same kind of institutional gatekeepers, economic barriers, or social barriers common in the news industry. In principle, anyone can add or edit anything to Wikipedia, including groups that are poorly represented elsewhere.

Because Wikipedia is often the first search result for topics, it has a deep influence on visibility of those topics in online media. How are women faring on Wikipedia? Women and men contribute to Wikipedia in different numbers. As of 2011, only 9% of Wikipedia editors were women, and they tended to make fewer edits than men (Wikipedia Editor Survey 2011). Women were much more likely to make edits about people or the arts than history or science. Women on Wikipedia didn't back down from contentious topics, although many of them leave when their edits are reverted. In individual cases, researchers found that women were no more likely to leave than men, but women's edits were reverted more frequently, leading to disproportionately high departures by women (Lam et al). The Wikipedia community has been aware of these biases since 2004, when the "Countering Systemic Bias" project began to address the imbalances resulting from Wikipedians' demographic tendencies.

Wikipedia has better coverage of men than women. The content of Wikipedia includes more people than the legacy encyclopedia Britannica, including women. A study of Gender Bias in Wikipedia and Britannica, Joseph Reagle and Laruen Rhue also found that Britannica is "more balanced in whom it neglects to cover than Wikipedia" (Reagle, "Nuance"). Comparing Wikipedia entries to a composite list across several encyclopedias and lists of notable women, Reagle and Rhue found that "while Wikipedia had nearly twice the number of female biographies than did Britannica, it had over two and a half times the number of male biographies" (Reagle, Rhue, 1145).
Wikipedia's biographies gender gap doesn't purely result from a lack of submissions about women. During the Smithsonia Women in Science Edit-A-Thon in March 2012, participants submitted new Wikipedia articles about women scientists. Although entries were developed with help of Smithsonian archivists, event organiser Sarah Stierch reported that several of the articles were nominated for deletion (Stierch). When the Wikipedia page for Kate Middleton's dress was nominated for deletion as not notable, Wikipedia Founder Jimmy Wales took the issue to the stage of the Wikimania conference as an illustration of the topic bias that can result from the predominance of male Wikipedia editors (Bosch). In Benkler's terms, while commons-based peer production does enable broader possibilities for self-definition, the demographics and structure of the peer group will influence and shape those possibilities.

### 2.5.5 Global Voices and citizen media

Given the evidence from mainstream media, feminist publications, and Wikipedia, it's easy to believe that analysis always leads to revelations of inequality. Yet some media organisations do have healthy levels of gender diversity. Consider for example, the international citizen media site Global Voices, where 51% of posts over its history have been written by women.

Global Voices is a global news site that translates and explains events and issues worldwide. A typical post curates tweets, citizen photos, and news articles for an international audience. Global Voices also translates and amplifies short summaries of content from regional sites, Although some editors are paid, most translators and writers are volunteers.

In an analysis with Irene Ros and Adam Hyland, we found that Global Voices improved and maintained diverse participation by women from 2005 to 2012, with women producing nearly half of all posts since 2007 (fig 4). All volunteer content communities exhibit a curve of participation, with a small number of people contributing most of the posts, so we calculated the gender breakdown at different numbers of posts per author. Contributions by women and men were nearly equal at all of those levels (fig 5). Finally,
we calculated the article gender diversity for several regions of particular note within the Global Voices community: Eastern & Central Europe, Latin America, Middle East - North Africa, and Sub-Saharan Africa. Women's writing is strong across all of these regions. Even in Sub-Saharan Africa, where the majority of posts are written by men, women are writing over 40% of the posts, which is far greater than the norm for many mainstream publications.

![Figure 4: Global Voices article author gender per year, 2005-2012](image)

![Figure 5: Global Voices author gender, grouped by posts per author, 2005-2012](image)

Irene, Adam, and I interviewed Solana Larsen, managing editor of Global Voices, to learn what makes Global Voices more diverse than other media organisations, even
though gender diversity was never one of its stated goals. According to Solana, Global Voices selects for people who "see the world in terms of shared experiences and similarities rather than differences." Global Voices editors and contributors share an interest in highlighting other people's voices. The founders, board, and editors of Global Voices are gender balanced. Because Global Voices focuses on social issues, Solana believes that its writers are very conscious of issues surrounding discrimination and equality. Finally, because diversity is the norm, gaps in gender representation are easier to notice, call out, and adjust.

Women have typically been poorly represented in mainstream media, and their role in American newsrooms is actually diminishing. Women who do make it into newsrooms find themselves focusing on entertainment news, and tactics to create change have fallen far short of their goals, sometimes even losing ground. Even online, where costs of publishing have reduced and the gatekeepers of voice have been restructured, women's writing is hard to fund, and gender imbalanced volunteer cultures can prevent women from being heard. Despite this gloomy picture, sites like Global Voices demonstrate that it's possible to foster diverse content cultures online.
3. REPRESENTATION IN NETWORKS

Representation is often studied in terms of content and reception. Content analysis reveals ways that a group is presented in articles and comments. Reception analysis examines the views and responses of those who encounter that content. Online, the study of representation and interventions for changing representation take into account the network of cultures, brands, and algorithms that participate in the voice and visibility of women in society. Modeling representation has become more complex as media power transforms from broadcast power to power in networks, with mainstream media brands participating in an evolving social, linked ecosystem of attention, information flow, and conversation. Technologies that address representation will acknowledge the structure of power in these networks, their influence on representation, points of intervention across those structures, and the critical role of data collection in informing those interventions.

3.1 Media Action for Representation

There are multiple ways to use the media to address issues of representation, theories of change that describe media actions and the hopes that accompany them. Advocacy uses the media to appeal to authority. Norms-creation is the act of packaging media to argue for the presence of particular social and political norms. Conversation-starting introduces an issue to an existing community. Convening power is the ability to influence who participates in a conversation. Cooperation is the act of facilitating groups of people to carry out coordinated tasks.

3.1.1 Advocacy

Advocacy campaigns set out to create change by appealing to authority. If an article or a set of petitions are shown to the right authorities from a respected news brand, authorities are expected to exercise power in response (Stempeck et al). Projects like Vida Women in Literary Arts and the Op Ed Project apply this tactic. In the Op Ed Project, individual women are mentored to use their voices in mainstream media outlets in the hope that they will gain other opportunities from powerful people who read those publications. VIDA publishes reports to pressure and convince literary publications to publish more reviews.
of women’s novels. Advocacy is also at work when a literary publication publishes a review; it is an appeal to consumers to purchase and read someone’s work.

3.1.2 Norms

Norms-creation has typically been understood as the ability of broadcast media to participate in the ongoing development of cultural norms by curating examples that convince people of that norm. In journalism, political norms are often discussed in relation to Hallin's spheres, a set of overlapping topic circles which shift over time. The "sphere of consensus" refers to coverage of issues on which everyone appears to agree. News organisations also address the "sphere of legitimate controversy," including a variety of voices which disagree. The "sphere of deviance" includes issues and people who are considered unpalatable or "unworthy of being heard" (Rosen).

We can imagine a parallel set of spheres for representation in broadcast media. If women are talked about but their voices are systematically excluded, they are in the sphere of deviance. If they are included as experts on some topics only, then they remain "other," included only in the space of legitimate controversy. When women are equal producers of media and not just subjects of media, a publication both presents and reflects a social norm for women's equal voices in society.

Spheres of consensus online can be established through curation on websites and social media in addition to packaging by broadcast organisations. If everyone in your Twitter feed or the blogs you follow is sharing similar links, you may come to accept those issues and voices as norms, even if they constitute a minority of coverage in their respective publications. Broadcast organisations need not be involved at all in this kind of norm shaping. If large numbers of your Facebook friends are changing their profile image red in support of marriage equality, as many did in March 2013, these norms become visible even in cases where mainstream media focuses on other issues.

3.1.3 Conversation-starting

Online, fostering and provoking substantial conversation has become an art of its own. Many conversations include strong disagreement with an article. Conversations about a
single article can take place across thousands of email lists, discussion sites, friendship clusters within social networks, and the comments of the originating publication. In many of those contexts, comments by individuals, especially friends of the reader, can carry equal or greater rhetorical weight as the piece of content itself, even when those individuals critique or disagree with the linked content. Tactics which call out sexist behavior and speech often start a conversation by re-posting material from one context into online spaces where discussion of gender and sexism are common.

3.1.4 Convening
Convening, the ability to influence who participates in a discussion, most often appears in the news when a journalist chooses sources; quotations in an article create a conversation among people who may never have spoken with each other. Online, we also exercise convening power when we create an email list, mentioning specific people in a post to Facebook or Twitter, or add and remove people from Branch conversations.

Counterpublics convene conversations away from the visibility of the broadest networks. In “Rethinking The Public Sphere,” Nancy Fraser argues for the recognition of alternative spaces where women have historically carried out politics and conversations outside the most visible areas of public discourse (61). These spaces have been a necessary response to women’s exclusion from power, and they also offer critical support for women to find peers, develop ideas, and advance their voices more visibly. Catherine Squires expands this notion by describing publics in the black public sphere in terms of their varied missions and varying relations to mainstream media. These enclaves, satellites, and counterpublics support underrepresented groups by convening as supportive alternative spaces where links to broader networks are only selectively accepted.

3.1.5 Coordination
Coordination is the art of directing the behaviour of participants toward common goals. Wikipedia and other commons-based peer production technologies support complex cultures of coordination. Other technologies for coordination include services that modify
a Twitter profile to express solidarity with a cause, website banners that direct all of your visitors to a call to action, and Twitter-bombing interfaces that coordinate participants to nag a celebrity to support a cause.

3.2 Networked power

In "A Network Theory of Power" Manuel Castells sets out to identify forms of power occurring in an environment where "no unified power elite is capable of keeping the programming and switching operations of all important networks under its control." This ecosystem, according to Castells, needs to be described in terms of the kinds of power that different actors carry out on the structure of networks and the flow of information across them. Within Castell's model, the two kinds of power relevant to networked representation are the power to share voices from one network to another and the power of networked actors upon each other.

3.2.1 Switching power

Cross-network "switching" power is exercised when a blogger is quoted in mainstream media, when a tweet from a frustrated citizen is discussed in an online community, or when a meme starts in a specific community and is remixed across the social web. The power of network actors on each other could be something simple like the social influence of our friend's preferences on our own attention habits. A more complex example might be the competitive content pile-up which sometimes occurs when news outlets, who monitor viewer behaviour, all decide that they need to publish an article about some timely topic in order to collect maximum advertising revenue from viewers.

Cross-network power often requires what Ethan Zuckerman calls "bridge figures," actors who are able to interpret between two networks by understanding media from one network and sharing it with another network in a format that can be received (Zuckerman, “Bridge blogger”). Zuckerman focuses on cultural bridges: people who can interpret and explain issues across language and culture. Global Voices, which Zuckerman co-founded, is an entire community established to facilitate cross-network
bridges. Coordinated with a combination of technologies and editorial practices, Global Voices contributors curate, explain, and translate conversations across geographies and cultures. Writers who share conversations from feminism with a general audience create bridges across networks as well.

### 3.2.2 The influence of network actors upon each other

In digital networks, the scale of conversation can be too large for comprehensive human intervention. For this reason, bridges can also be pieces of software which observe speech in one network and automatically translate the digital format to speech which is compatible with the technical and social codes of another network. These automated bridges often amplify the second kind of power Castells’ discusses: that of networked actors upon each other.

Power in networks is often mediated by algorithms which blend the actions of individuals and organisations with broader patterns of behaviour and formalised social codes to decide who gets heard by whom. When you see an article on Google News, your awareness of the headline is mediated by the Google News recommendation algorithm, which processes the contents of thousands of media outlets and decides which ones to show. When I see conversation about an article you posted to Facebook, the News Feed algorithm interprets what to do with your act of posting, in the context of our relationship history, the number of comments and likes it has received from our mutual friends, and the formal codes each of us sets to define our personally-defined privacy and censorship settings.

### 3.2.3 Data in networked power

Those who possess data about online interactions and the power to surface that information have strategic advantages to exercise power in the media. The first of these advantages is the knowledge about where power lies in networks at a given time. For example, the application of advocacy power requires knowledge about ways to attract the attention of the people to be influenced. If you know that 52% of links which reach the Reddit front page have been posted to the site more than once, you can adjust your
behavior on the platform to increase your chances of getting your links onto the front page (Gilbert). Newsrooms guard traffic data carefully because access to that data can afford strategic knowledge about reaching their audience.

Data on digital network interactions also affords large-scale testing and evaluation. Experiments and data collection can validate hypotheses about the behavior of network actors upon each other, as well as the effectiveness of strategies for bridging and switching. These experiments can be carried out with tens or hundreds of thousands of participants at a high rate of iteration and improvement.

3.3 Representation in Networks
Understanding networks is critical to understanding representation in contemporary media, where visibility occurs through interactions across networks. Networked representation describes the contours of visibility that a particular group has across media networks. In these networks, the question of whose voices are heard by whom online is mediated by networked gatekeepers that are human, institutional, and computational. Data about network interactions affords strategic power to those who can access and analyze that data, especially for evaluating and improving interventions at a high rate of speed over very large sample sizes.
4. MEASURING GENDER REPRESENTATION IN CONTENT

Monitoring media representation online can take advantage of high speed automated data processing and analysis. Automated techniques can replicate some kinds of manual content analysis, including the gender of who’s speaking, the gender of content subjects, and the gendered nature of language used to describe news subjects. These automated techniques offer greater detail, breadth, speed, and focus than manual approaches to content analysis.

4.1 Manual Content Analysis: Global Media Monitoring Project

Content analysis is the primary source of evidence on the representation of women in media. The Global Media Monitoring Project has conducted an international analysis of women in the news since 1995, when they started with 71 countries. Every five years since, they pick one day of the year and tabulate information on the gender of who was speaking, the topics where articles refer to and quote women, the language that is used to refer to them, and the images used to portray women across print, broadcast, and online news sources. Repeated most recently in 2010 with data and case studies from 108 countries the report documents global evidence for the systematic exclusion of women in the media,

The Global Media Monitoring project is one of the most sophisticated examples of cooperative media analysis centered on manual coding of content. Journalism students and other volunteer "coders" from around the world collect a day of news from 1,281 newspapers, television shows, radio stations, and an additional 76 news websites. Looking through each of those newspapers, students tabulated data on the speakers, language, and topics of 16,734 news items, 20769 news personnel, and 35,543 news subjects. Individual items are analyzed by multiple coders so that tabulated data from individual coders can be compared for inter-coder reliability and aggregated into a final report. If discrepancies are found across coders, whole sections may need to be re-coded, as happened with Spanish language media analysis for the 2010 report. The process starts with partnership building, the development of transnational taxonomies, and the creation
of volunteer guides in multiple languages. It takes years to complete coding, aggregation of data, verification, and the creation of a final report.

Human coding of news is especially effective for tasks which require nuanced interpretation. For example, the Global Media Monitoring project tracks how often women are referred to in terms of their occupation rather than their marital status across 26 categories ranging from lawyers and coaches to suspects and parents. Human coders categorize whether women speakers are offering popular opinions, opinions based on experience, eyewitness reports, commentators, spokespersons, or the subjects of stories themselves.

![Figure 6: Functions of news subjects by sex, by occupation: 2010, Global Media Monitoring Project](image-url)
Content analysis projects like the Global Media Monitoring Project have two theories of change beyond the reports they publish, which are themselves helpful to make sense of the role of women in the media. Content analysis projects foster networks of people and organisations passionate about supporting women's voices in the media, networks which spread lessons and ideas for creating change. Secondly, the act of content analysis itself is seen as a positive learning experience for the journalism students who participate. If writers are trained to be attentive to others' writing, perhaps they will be more attentive to their representation of women in their own writing.

4.2 Cooperative Online Content Analysis: PageOneX

The speed and scale of content analysis by human coders can be increased by online, collaborative crowdsourcing technologies. The front-page analysis software PageOneX supports multiple volunteers who can highlight parts of newspaper front pages that feature a topic of interest. The software coaches coders on identifying topics, and advises them how to tag their highlights. PageOneX automatically calculates the area of pages dedicated to a particular topic, and calculates intercoder reliability, automating the process of verifying results and aggregating them into a report.
The above example illustrates the degree to which online software can increase the efficiency of human coding. Pablo Rey Mazon used PageOneX to classify the percentage of front page surface area given to articles by men, women, and mixed gender collaborations. In less than an hour, Pablo was able to code two a week of front page news across two newspapers, while also auto-generating a visually compelling report.

4.3 Automated Content Analysis

Automated methods of content analysis can enable the analysis of gender in media ecosystems across millions of records in near-realtime. Although they offer less nuance than human-coded analyses, automated methods are a necessary building block for the analysis of representation in networks. Across my thesis, I apply simple techniques for measuring the gender of authors, subjects, descriptive language, and people quoted in text.
4.3.1 Name Gender
The gender of who's speaking can be estimated by extracting the first name of the speaker and matching the first name with demographic data on the probability of that name being male or female. The open Global Name Dataset, compiled with Irene Ros and Adam Hyland, features name statistics for the United States, England and Wales, Northern Ireland, and Scotland. The US name gender dataset incorporates all first names since 1938 with a minimum incidence of 5 births per year, along with spottier records going back to 1880, for a total of 89925 names. The United Kingdom data, from the Office of National Statistics, incorporates full name data from 1996 to 2011, except Scotland, where name data has only been recorded for 2009 and 2010. The UK dataset includes 30715 names, although all of the UK analyses done here are with a smaller ONS dataset of 27548 male and female names. With a few exceptions like "Pat", "Abba", and "Nicky", many names have a high gender probability. When name probabilities are matched with author names, it's possible to calculate author gender very quickly across millions of articles.

4.3.2 Subject Gender
Overall gender of the subject and speakers within text can often be estimated by counting gendered pronouns. Quotations often end with the phrase "he said" and summaries often use gendered pronouns to talk about the actions of people discussed in an article. Pronoun counts were used to classify the gender of biography entries in Joseph Reagle and Lauren Rhue's study of gender bias in Wikipedia and Brittanica.

4.3.3 Language Styles
In some cases, it's possible to analyse the language used to describe men and women who are mentioned in text. At the Women of the World Hack Day, Michelle Brook and a team of developers built a system which created word clouds of the adjectives used by men and women journalists in the Daily Mail to describe men and women whose names appeared in the news. Named entities were extracted using OpenCalais and adjectives within those sentences were counted to produce a frequency distributions which informed the word cloud (Brook).
4.4 Characteristics of Automated Content Analysis

Automated analysis of content can include every piece of content (called populations) rather than the limited samples of projects like the Global Media Monitoring Project, which looks at one day every five years. Analysis of population data can be helpful in identifying overall trends, since the content of news can vary dramatically from day to day. Even when study designers work carefully to pick a day likely to be “representative,” i.e., free of unusual news stories that might skew coverage, they are likely unable to account for gender variation over time.

4.4.1 Variation over time

Consider, for example, this chart of content gender in the New York Times from 1987 to 2007, estimated with pronoun counts (LDC). As you can see, the percentage of articles per month about men and women repeatedly varies by more than 20% within a single year. It's possible that some human coded studies are hinting at false trends, especially for parts of the media where the content gender varies widely in the news.
4.4.2 Interpreting automated content analysis in context

As we saw with book reviews, opinion pages, and Wikipedia entries, large-scale analysis of content over time still needs to be tailored to the culture and practice of the content being analyzed. Consider, for example, obituaries in the New York Times. Obituaries offer a public remembrance of people whose lives are considered notable. Unlike articles that focus on a single event, obituaries present the narrative of an entire life, describing the circumstances of education, opportunity, decision, and relationships that combine in the life of a notable artist, scientist, entrepreneur, or public servant. Obituaries present a perspective on the social standing of women in society and establish a vision of what’s possible in life.

The absence of women in obituaries creates a gap in the collective vision of what women can achieve. The average percentage of obituaries about women and men in the New York Times remained fairly stable from 1987 to 2007, only occasionally moving beyond 20% of obituaries per month.
Splitting editorial obituaries from paid obituaries reveals a more dynamic story. The number of editorial obituaries about men and women reduced dramatically in that period, marking a shift from more local obituaries to offering more detail about a smaller number of internationally-known figures. This is in contrast with paid death notices, which have a very different editorial process, exhibit less of a gender disparity, and which have remained more stable in numbers. Even within the same paper, the market forces of paid death notices produce very different gender representation than the editorial decisions applied to obituaries. The following charts are screenshots from interactive visualization sketches by my collaborator Sophie Diehl.

4.4.3 Filtering and focused exploration

Computational analysis of gender across every article of a public permit focus as well as breadth by offering the ability to filter content by person and topic. In an interactive visualisation sketch I produced in search of a story for the Guardian Datablog, I charted
the most prolific Guardian authors per week from June 2011 to July 2012. Had I chosen to conduct further analysis, I might have created something akin to NYTWrites, an interactive visualization by Irene Ros which illustrates authorship patterns by sharing which writers are covering what topics across New York Times content. Rather than simply presenting datapoints, interactive reports like this invite participants to explore focused subsets of data within issues of media representation. The following interactive slopegraph that I created shows a weekly ranking of the most prolific and most popular writers in the Guardian. When the user selects an entry, the chart highlights that writer’s ranking for other weeks.

Figure 13: Slopegraph chart of most prolific Guardian bylines, July 2011 - June 2012
Computational analysis of gender representation also changes the rate of reporting from months or years to seconds. Post-hoc studies like my analysis of the New York Times with Sophie Diehl or Reagle and Rhue's analysis of Wikipedia focus on a specific period of time. These studies can take weeks or months to set up and conduct and get quickly out of date. In contrast, Open Gender Tracker is a technology developed together with Irene Ros and Adam Hyland of Bocoup which accesses data of the latest content from newspaper APIs, producing gender reports on that content, and sharing it back in machine readable form. In one of two Open Gender Tracker demos, we produced a live website which can show faceted timeseries of Boston Globe content gender up to the latest data available on the API.

Open Gender Tracker also offers flexible, exploratory access to gender representation data, since it attaches gender information to the whole range of possible queries to the
Boston Globe API. If a user requests that a report be narrowed by date range or facet, Open Gender Tracker passes the request on to the requisite API, processes the results, and presents the results with relevant gender information added. Those results are cached in case the request is repeated.

Computational technologies for content analysis like Open Gender Tracker can produce machine readable output to be read by other software systems. It can share data with the statistical software R to support detailed analyses conducted by researchers. On the web, Open Gender Tracker can also serve results in JSON format, easily accessible by the network of online algorithms which share information and take action on the social web.
5. MEASURING REPRESENTATION IN NETWORKS

Since the web makes audience activity countable, integrating it into the machinery of social conversations and information flow, both research and action on representation must integrate audience activity with content analysis. It also opens new stances for news organisations, supporting a shift from defensive responses to critiques of misrepresentation to a search for markets receptive to the fair representation of women. The most comprehensive, strategic analysis, I argue, moves one step beyond audience analysis to measure the curation of information from content to conversations and back into content.

5.1 Attention as representation

Audience attention for women's writing can vary substantially by publication. In a study of a year's UK news, Lisa Evans and I found very different preferences for women's writing between the the Guardian, Telegraph, and Daily Mail newspapers. Readers of the Guardian and Daily Mail tend to share a greater percentage of women's writing than the proportion of what the newspapers publish. Guardian opinion articles by women are 35% of what gets shared, compared to 30% of what is published. Only 21% of the Daily Mail's opinion articles are by women but those women generated 35% of the opinion section's shares and likes. In contrast, Telegraph readers don't share articles by women as much as the other two papers. While women write 20% of its opinion articles, their articles are only 13% of what Telegraph readers choose to share (fig 15).
5.1.1 Measuring social media attention

To draw these findings, we wrote software to download every article published by each of the three newspapers between July 2011 through the end of June 2012. Guardian articles were accessed through the Guardian Open Platform API, which publishes machine readable data about the Guardian's content, including titles, bylines, categories, and public web addresses for each article. Daily Mail and Telegraph articles were downloaded from the public websites of these two publications, by writing software to browse their content indexes, identify article addresses, download individual articles, and extract title, content, and byline information from individual webpages. Dates were obtained through the content indices, and categories were inferred from the structure of the web addresses, which include a taxonomic information. The full collection included 76,644 Daily Mail articles, 143,515 Guardian articles, and 110029 Telegraph articles. Social Media likes and shares were attached to each article by sending the article URL to the public APIs of Facebook, Google+, and Twitter, during the first week of August 2012. Those services returned counts of how many likes, shares, and +1s were received by articles within that period. By joining social media records with article and category records, we were able to create an interactive visualisation of social media activity per section per week across each of these papers.
These findings illustrate the importance of including audience data into measurements of representation, even if they cannot offer answers on the reasons for Telegraph audiences' disproportionate sharing of women's writing or the specific nature of the content shared. The data collected also offers a window into approaches which could answer those questions. More thorough analysis of women's opinion writing in the UK could examine the relative popularity of the topics women write about, compare women's article popularity per topic to the norm, or compare the social reach of individual writers to each other.

5.1.2 Measuring reader demographics

Audience metrics companies like Quantcast use a combination of surveys with tracking technologies to estimate the demographics of visitors. For example, their April 2 report on The Telegraph claims that readers are more male than average, more affluent, have university educations, and tend to have no children in the household.
5.1.3 Measuring referrer data

Referrer data collected by webservers offer another source of data about audience behaviour across networks. In an article on what he calls "Dark Social," Alexis Madrigal of The Atlantic explains information collected by content sites when visitors access a page. When a browser visits a website, it often shares information about the web-page previously viewed. Using this data, it is possible for publishers to distinguish the source of much of the traffic referred from elsewhere on the web across a diversity of platforms, content styles, and conversational cultures, including the search terms used to find content.

5.1.4 The objectives of private and shared network data

The analysis of representation only partially overlaps with the goals and evidence of content producers. Publishers use data to evaluate and prioritise the channels they use to spread their content, maintain their brand, and attract viewers to their advertisers. When Martin Belam calculates that UK newspapers have lost 27.4% in print in the last five years, the Guardian has lost 40% of its print subscriptions, and that The Guardian's Facebook traffic is passing search in volume, he's using that data to comment on The Guardian's publishing priorities (Belam, “Seismic”; Belam, “How”). The computational analysis of representation on the other hand is fundamentally comparative and often conducted on the outside of publications, without access to detailed traffic data. Social media metrics allow a more nuanced picture of representation than simple counts of articles. Rather than tracking how much content is available, they offer signals that allow us to estimate the readership for that content.

5.2 Curated Representation

Content curators are another major part of the network of representation online. On Madrigal's chart of sources of traffic to The Atlantic, the combined traffic from Reddit, Hacker News, Digg, and StumbleUpon exceeds traffic identifiable tracked to Twitter. With the exception of Digg, these content curation sites invite users to post links for others to vote on. As third party links receive more votes, they rise in prominence within
the social curation site. Reddit, Hacker News, and Stumbleupon all offer APIs which serve data about users voting behaviour for content posted to their sites.

5.2.1 Automated representation
Data from social media and content curation sites participate in a feedback loop in which algorithms observe and influence audience attention towards particular links. This is most visible on the website Digg, which relies entirely on signals from other sites like Facebook, Twitter, the link sharing service bit.ly, and the content analytics platform Chartbeat(Van Grove). As print and web front pages continue to drop in influence, the role of editors in shaping women's representation is being replaced by these curatorial algorithms, observers of behaviour across multiple social communities throughout the web.

5.2.2 “Viral” representation
The spread of information on social media always happens in conversation with curators, brands, and algorithms, and almost never spreads "virally" from person to person. In a comparative study of multiple games, microblogging services, and communication platforms, Sharad Goel and others from Yahoo discovered that "adoptions resulting from chains of referrals are extremely rare" and that "the bulk of adoptions often takes place within one degree of a few dominant individuals." Although networks online enable us to share things with our immediate circles, things which are seen widely are still often channeled through those with large audiences.

5.2.3 Feedback between citizen media and mainstream media
It would be a mistake to imagine a two-tier content machine where legacy and new media brands broadcast content whose relative popularity is defined by a feedback loop with audience behaviour. Content originating outside of media brands often gets featured by the mainstream media when it becomes popular enough to attract commentary or advertisement. Reporting practices now routinely involve quoting voices from citizen media, and new brands can arise unexpectedly from the attention associated with a cultural moment.
The interactions between content providers can be measured by extracting hyperlinks found within online content. In a talk at the Ford Foundation, Ethan Zuckerman illustrated this by talking about interactions between political websites and mainstream media coverage of the shooting of Trayvon Martin in early 2012. Media Cloud researchers at the Center for Civic Media and Berkman Center counted which sites received the most links from the mainstream media to estimate which blogs they thought were having the most influence on content appearing in mainstream media. According to Zuckerman, blogs and thinktanks with the greatest number of links can be said to have the greatest influence on the representation of a person or issue in the media.

5.3 Interactions between publishers and online networks

As journalists increasingly turn to social media as a source for photos, footage, and commentary, it's also possible to track the representation of individual voices across content networks. Citizen journalism and social media have become popular sources for the news after the Arab uprisings of early 2011, where social media was often a primary source for breaking news. During the Arab uprisings, a small number of figures became very prominent sources for the media, raising questions about whether citizen media sourcing was increasing or decreasing the diversity of voices in the news (Lotan et al, “Revolutions”). Questions of source diversity in social media citation can be answered by extracting citations of social media from news content.

5.3.1 Measuring social media sourcing in Global Voices

In my initial exploration of Twitter citation in the news, I processed articles from nine years of Global Voices content, extracting data for every tweet and twitter account cited in the content of posts and storing information about twitter accounts that appear in the same post. I used that data to create an interactive visualisation of Twitter citation in Global Voices over time, faceted by region, with additional information about content headlines and comparative rankings of source popularity.
In the case of Egypt, Global Voices featured tweets since 2005. The first major spike in blog coverage occurred in February 2007 when blogger Kareem Amer was sentenced to prison for statements made on his blog. The next spike in coverage, in February 2009, occurred in response to the Cairo bombing, which is also correlated to the first substantial use of Twitter as a source in the Egypt section. The largest spike in Egypt coverage starts at the end of January 2011 in response to protests in Tahrir Square and is sustained over the next few weeks. Notice that while Global Voices did quote Twitter from time to time (citing 68 unique Twitter accounts the week of the Cairo bombing), the diversity of Twitter citation grew dramatically during the Egyptian uprising -- and actually remained consistently higher thereafter.

Whose voices was GlobalVoices quoting? Citation in blogs and the news can give a source exposure, credibility, and a growing audience, since readers can click on a person’s name to follow other things they are saying. In the Egypt section, the most cited Twitter source was Alaa Abd El Fattah, an Egyptian blogger, software developer, and activist. One of the last times he was cited in Global Voices was in reference to his month-long imprisonment in November 2011.
Although Alaa is prominent, Global Voices relied on hundreds of other sources. The Egypt section cites 1,646 Twitter accounts, and @alaa himself appears alongside 368 other accounts. One of those accounts is that of Sultan al-Qassemi, who lives in Sharjah in the UAE, and who translated Arabic Tweets into English throughout the Arab uprisings. @sultanalqassemi is the fourth most cited account in Global Voices Egypt, which accounts for only 28 posts out of the 65 where he is mentioned. Al Qassemi is quoted across a more diverse range of topics than Alaa, who is cited primarily just within the Egypt section.
Encouraged by these results with a single publication, Diyang Tang and I built a dashboard which shows who from Twitter has been quoted across the news within the previous day. Processing data from thousands of international media sources available through the Media Cloud API, we collected data on the quotation history of every Twitter account cited up for months, up to the current day. With this constant flow of recent information about who from social media is being quoted in blogs and the news, it becomes possible to aggregate data about the gender diversity of social media sources in news and blog content at any moment in time.
5.3.1 Broadcast and pop-up brands on social media

Social Media metrics of the news, content analysis, and source diversity measurements assume that the brands and platforms of media ecosystems are stable— that it's possible to point a data collection system at twenty thousand websites and approximate the state of information flows. The story of “Binders Full of Women” shows that unexpected interactions of news, PR, and broadcast media can lead to the rise and fall of powerful conversations in response to interest that hasn't yet found an object of attention. When that interest is colonized soon enough and audience attention is retained, pop-up brands like the Binders Full of Women meme can become an effective longer-term platforms for reaching audiences with voices that aren't well represented in the mainstream media.

When Mitt Romney talked in the second televised US presidential debate of 2012 about the "binders full of women" he used to select women staff members, the phrase became a catalyst for online critiques about Romney's attitudes for women. With 65.6 million viewers watching, the potential was already high for politicians’ phrases to become popular Internet users made jokes, created parody meme images, and left thousands of satirical comments in Amazon reviews of binders. Twitter accounts, a Tumblr pages, and
a Facebook page curated people's creations. Although media attention focused on
Veronica De Souza, creator of a Tumblr page which received 11,000 followers in the first
few hours, a Facebook page for Binders Full of Women reached 100,000 subscribers by
the end of the debate. De Souza's Tumblblog, like many one-off Internet jokes, concluded
the day after the presidential election, less than a month after the debate (Stenovec). The
Facebook page however continues to be active and facilitates a lively ongoing
conversation about women, culture, and US politics.

The creator of the Binders Full of Women Facebook page didn’t see himself as an activist
or media creator before deciding to create the page. “The page started on a whim,” he
stated in an email interview. After seeing Romney’s statement in the debate, “I opened by
Facebook tab and searched ‘Binders Full of Women.’ When nothing came up, I decided
to make a parody page, never expecting much.” Already a self-professed “news junkie,”
the act of curating a Facebook page about “women's reproductive rights, income equality,
white privilege, institutionalized racism/sexism, and even gun rights.” The anonymous
creator of the page, reported a preference to his own content rather than accepting content
that advocacy organizations sometimes submitted.

Facebook makes detailed audience metrics available to creators of pages like Binders Full
of Women. The most basic reports offer daily information about the region, age, and
gender distribution of people who like and share content. More detailed data offers
information on how many stories appeared in the news feeds of other users after a
particular post was liked by someone who follows the page, along with timestamps for
each like, comment, and share.

5.4 Measuring representation in networks
Computational analysis of representation of networks, when combined with content and
and audience analysis, offers the possibility of tracking women’s representation online.
Much of this data is held closely by individual pages and publishers. Despite this,
publicly available data can be used to measure attention, curation, and the sourcing
practices that carry out interactions between publishers and online networks. Although
analysis of established publishers and networks can miss the rapid rise and fall of pop-up brands, the available metrics provide a rich resource for data-driven tactics to improve women’s representation in the media at key points of influence.
6. POINTS OF INFLUENCE AND THEORIES OF CHANGE

Theories of change around women’s representation tend to focus on mainstream publishers. These theories argue that fair representation could be achieved if newspapers agreed to hire, include, and write about women using fair language. However, now that more people's voices are visible online, it’s more complicated. New tactics of representation focus on spreading individual messages, media transparency, confrontation, and public shaming online. These approaches tend to focus on specific cases or publications rather than address systemic biases. I argue that computational tactics for women's representation can address those broader systems, offering tactics to monitor and address systemic biases over time.

6.1 Short-term advertising campaigns

Advertising approaches to influencing networks focus on reaching a large number of people with carefully crafted pieces of media, or establishing an ongoing relationship with fans that appreciate a certain type of content. Campaigns like KONY2012 used a combination of social media and other networks to widely spread a single video (Lotan, “KONY”). Companies like UPworthy extensively test the spreadability of each political video and image they share, using the popularity of each piece of content to grow the overall reach of their brand. Despite the temptation to create a campaign to create audiences for women’s voices, this theory of change is likely to draw audiences to a small number of publications, not create change on a widespread basis.

6.2 Making representation visible through transparency

One straightforward approach to changing journalism with computational content analysis is through transparency: if data on women's voices is made visible to news producers and consumers, perhaps the people and organisations involved will change. In the US and UK, the Media Standard Trust and Sunlight Foundation's Churnalism project attempts to create this kind of transparency in the area of press releases and original journalism. Confrontational tactics call out misrepresentation publicly to pressure organisations or communities to change their practices.
6.2.1 Churnalism: automated accountability and skepticism

Applying software similar to plagiarism-detection systems, Churnalism tries to detect cases where newspaper articles quote press releases or other articles. The US website invites readers to paste web addresses or text which they suspect might be recycled content. The Churnalism site highlights other sources which share similar text. "Discover the journalism you can trust and what you should question," the website reads.

The Sunlight Foundation carries out two parallel theories of change. By exposing the transactions of power, they set out to foster public skepticism about those who exercise power. Secondly, their data is sometimes the basis of accountability journalism. By collecting data on transactions of power which the Sunlight Foundation sees as ethically dubious, they provide evidence to those who set out to challenge that power. These challenges often take the form of accountability journalism.

6.2.2 Confrontation about women's representation

Transparency isn't working in the case of gender representation. Confrontational appeals to authority have largely failed for a half century. Organisations like VIDA report that publishers often entrench their position when confronted with data on women's voices in their content. It’s not clear that faster, more efficient, more targeted confrontation will necessarily foster healthy, diverse representation online.

Occasionally, collective confrontation about the representation of women does move publishers to respond to isolated cases. On March 30, a New York Times obituary about rocket scientist Yvonne Brill emphasized her domestic life instead of leading with her substantial engineering achievements. The article opened with the statements that "she made a mean beef stroganoff [and] followed her husband from job to job" (Sullivan). After readers responded angrily on Twitter, the New York Times adjusted the article text. This small change was a token victory, since women remain a small minority of people celebrated in society, including the New York Times obituary section.
6.2.3 Nohomophobes.com: automated transparency of hate speech

Content analysis software has the capacity for automated transparency and confrontation about gender representation, moving beyond prominent individual cases to challenge widespread norms. The website No Homophones automatically calls out Twitter users who use terms like "faggot," "no homo," and "dyke" in Tweets in realtime, linking viewers to the Twitter account of those users and aggregating incidences of those terms by day, week, and over time. "Speak out when you see homophobic or transphobic language from friends, at school, in the locker room, at work, or online," it urges. Although small print at the bottom of the site claims that the site makes no claims about the homophobia of the user, the site prominently displays the names and photographs of many of those users on its front page.
Technologies like No Homophobes, which make the speakers of objectionable speech more visible, participate in what danah boyd calls radical transparency, "the idea that forcing people into the open will force them to behave civilly." boyd is responding to discussions about the requirement to use real names on social media platforms like
Facebook and Google Plus. These companies argue that people can be held more accountable for bullying and online cruelty if their identities are publicly associated with their actions online. boyd argues that attention-driven disapproval isn't applied equally, and that it destroys the lives of the most vulnerable more easily than people with power.

Radical transparency of speech online also risks misrepresenting the nature of structural injustices by focusing on the parts of systems that are most visible and easiest to track. For example, boyd argues that fears about online bullying sometimes obscure problems in schools because online speech is easier to track. Data on hate speech online contributes to the belief "that children today are at more risk than ever before even though, by almost every statistical measure, youth are safer today than at any previous point in history." boyd worries that out of proportion fear about hate speech online provides a rationale for censorship technologies.

6.2.4 Volatile outcomes of public confrontation on gender online

Confrontations in social media over claims of sexism can lead to bullying and loss of employment for both the accused and those who call out sexism in public. In March of 2013, Adria Richards tweeted a picture of two men at a technology conference who she overheard making sexual jokes. She directed the message to the organisers of the conference and to everyone else following the conference on Twitter, using the #pycon hashtag. The PyCon organisers talked to the individuals and dealt with the issue using their established code of conduct. When one of the men's employers announced that they had fired him, commenters online directed bullying speech towards Richards and her employer fired her.
For speaking up about sexism online, Richards faced "death threats, rape threats, a flood of racist and sexually violent speech, a DDOS attack on her employer -- and a photoshopped picture of a naked, bound, decapitated woman," points out Alice Marwick in an article about mob justice online in WIRED magazine. Marwick argues that the visibility of confrontation around sexism online has created ongoing battle-lines between feminist activists and "anti-misandrists," each of which justifies their actions by referencing the other side's actions:

While feminists believe it's important to call out people for sexist remarks to address structural gender inequality, another group believes calling out sexist remarks is just another example of women exaggerating harm [and] censoring reasonable behaviour''

(Marwick)

The destructive and hateful backlash against feminist speech online should lead us to question any belief in the inherent benevolence of transparency or the power of online mobs to police gender inequality in the media, especially in cases where transparency misdirects us away from the structure of women's misrepresentation to individual cases.

In the short term, the PyCon issue was addressed. In the medium term, it destroyed the career of the woman who reported it. In the longer term, this incident may well have a chilling effect on women's likelihood to report sexism in the technology industry.

In this thesis, I have chosen to avoid confrontational designs, despite the many requests I have received to create automated engines for public shame. Automated skepticism, confrontation, and shaming can indeed be made easier with content analysis technologies and spread widely on online networks. I acknowledge that confrontational tactics can create powerful solidarity around an individual story and sometimes lead to localised change. Yet these tactics have limited effectiveness at addressing structural biases, and the potential backlash may put at risk the very people they intend to support. Instead, I have chosen to imagine and create software interventions for systematic bias.
6.3 Software interventions for women’s representation in the news

This thesis presents three examples of alternative interventions to support fairer representation of women in the news, beyond advertisement and confrontation. Each of them accounts for sensitive questions of ethics and privacy. Although their effectiveness can only be determined after broader deployment, each project offers a clear theory of change and the means to evaluate its effectiveness toward that change.

6.3.1 Gender metrics platforms

Automated content analysis software can help sites monitor their diversity and impact. The diversity of Global Voices and the gender bias problems of Wikipedia illustrate the importance of collecting diversity data for volunteer sites. Community editors could review that data to evaluate the state of diversity on their site and take appropriate responses.

Counterpublic media in the feminist blogosphere and mentoring programs can use content tracking systems to evaluate impact beyond the walls of their publications. One can imagine a system which tracks the bylines of people trained by the Op Ed Project, as their writing spreads across multiple brands and their following on social media grows. Just as the Twitter quotation dashboard tracks quotations across tens of thousands of publications, similar software could tracks who's speaking across that same dataset.

Automated gender-tracking systems for news businesses could help news organisations find and keep diverse audiences for diverse content. Open Gender Tracker adds gender metrics to information from newsfeeds and news APIs. It offers a first step toward systems that can link content with audience diversity, across who's speaking, who's quoted, and who's voices are shared.

6.3.2 Personal tracking software

FollowBias, a personal tracking technology, offers timely personal feedback on the gender ratio of a Twitter user’s actions online. A complete version could help readers use measure their own media consumption, and journalists could use it to monitor the
language they use when speaking about women, or whether they speak about women at all. Content curators could track the gender ratio of who they're retweeting, quoting, and sharing to their audience. Personal trackers deliver positive norms for diversity, give users flexibility in their interpretation of those norms, and offer simple metrics for personal change. The data tracked by users can be aggregated to answer the question of impact: does access to such a technology lead users to change the gender ratio of their activity over time?

It's unrealistic to expect everyone to use a personal tracker for media consumption. It's much more plausible to design personal trackers for journalists and curators, whose choices to read, quote, and share voices have a greater impact on women's representation across networks. For this reason, the initial version of FollowBias tracks the gender ratio of who you follow on Twitter. Further designs could focus on key actions in other platforms. For example, a personal tracker for conversation site Branch could track the gender ratio of who users invite into conversations.

6.3.3 Participatory platforms to address contributor and content disparities
I believe that automated media monitoring campaigns like Churnalism and Nohomophobes.com encourage a kind of dismay and cynicism that doesn’t support ongoing, constructive change. “Passing On,” which I created with Sophie Diehl, directs readers toward constructive action using data on the lifetime achievements of women celebrated in mainstream media. The project visualizes obituary gender over 20 years of the New York Times and invites readers to use that data to check and improve Wikipedia’s coverage of women. In Passing On, any action from reading a page to checking a source contributes to a cooperative project to respond to bias in the New York Times by improving Wikipedia. Collectively, these projects represent a possible approach to issues of longstanding systemic bias that hope to nudge individuals towards sustainable change, rather than naming and shaming bad actors, which I do not believe can effectively address broader disparities.
7. OPEN GENDER TRACKER, A PLATFORM FOR CONTENT GENDER METRICS

Open Gender Tracker (OGT) makes automated, up-to-date gender analysis of content available to content publishers, media monitors, and organisations that support women’s voices in the media. OGT takes input from content APIs and datasets and produces reporting data that can be chained and combined with other reporting systems. Its open, extensible architecture is designed to fit easily with publishers' existing content platforms while support a flexible range of applications.

The OGT software was designed and built by Irene Ros and Adam Hyland at the open source company Bocoup, funded by a Prototype Fund grant from the Knight Foundation. It draws inspiration from software I created for analysis of the New York Times, Guardian, Daily Mail, and Telegraph. I advised its design and led case studies on Global Voices and the Boston Globe API.

7.1 Design
OGT uses a batch processing pipeline to append gender data to the data it processes. Job objects are collections of Article objects. Articles are created by Parsers, which convert external data into a format that OGT can read. OGT takes jobs from a queue, uses a Decomposer to prepare the data for the analysis, sends the decomposed content into Metrics objects, which then pass metrics to Aggregators that share the results.
OGT can be expanded by adding new kinds of parsers, metrics, and aggregators. The initial version of OGT includes parsers for processing content data from Wordpres sites and the Boston Globe API. Metrics currently can estimate content author gender and content subject gender. Output aggregators include JSON and CSV data for loading into spreadsheets and statistical software. Each of these can be extended to expand the capabilities of Open Gender Tracker. Expansion to other data sources simply requires the creation of a new parser type. Likewise, new OGT metrics, such as quote extraction or social media metrics data, can be added by creating new metrics. Integration a news organisation's data analytics platform would require the creation of a new aggregator.

7.2 Global Name Gender Database
To support development of further open technologies for gender analysis, Open Gender Tracker also hosts the Global Name Gender Database project, an open data collection of name gender statistics based on government data from multiple countries. Datasets for new cultures may be submitted via Github and distributed to any project that incorporate OGT. The dataset currently includes name gender statistics for the United States, England, Wales, Scotland, and Northern Ireland. It also includes scripts for adding gender analysis to the statistical software R. New techniques can also be tested against an open dataset of Global Voices content licensed under Creative Commons.
OGT's United States dataset includes all names between 1885 and 2001 with a minimum incidence of 5 births, a total of 89926 names. UK data from the Office of National Statistics includes all names with an incidence rate of greater than 3 across 1996-2011 for England and Wales, 1997-2011 for Northern Ireland, and 2009-2010 for Scotland, a total of 20,715 names. The Global Name Dataset was collected after all of the projects presented in this thesis, which use a smaller dataset from the US Social Security Administration's top 1000 names per year and the Office of National Statistics 15-year England and Wales dataset.

7.3 Open Gender Tracker case studies

Open Gender Tracker has been used in case studies analysing content gender in the citizen media site Global Voices and the content of The Boston Globe, a US newspaper. The previously discussed analysis of Global Voices, described, is an example of statistical exploration of a website archive using typical data analysis software. A visualisation app for Boston Globe content demonstrates OGT's ability to piggyback onto content APIs to create live data visualizations.

7.3.1 Processing archival data: Global Voices gender participation

Global Voices, a multi-lingual news site that curates international citizen media, has a reputation for gender diversity. In this case study, Global Voices bylines were analyzed with OGT to determine if the site was as gender diverse as its editors believed. The first research question investigated the gender diversity of Global Voices at all levels of participation, considering the possibility that the site's diversity may only have been among its most prolific authors. The second research question investigated gender diversity across regions. Followup conversations informed by that data investigated possible patterns of success for other content communities.

Data for this case study included all of the English language content on Global Voices from 2005 through 2012, exported in XML form from the site's Wordpress content management system. The Global Voices archive was comprised of XML files containing dates, topics, comments, and information on authors and translators. This data, which is
licensed under Creative Commons, has been archived publicly as a sample dataset on the Open Gender Tracker github page, stripped of email addresses and other sensitive personal information that is not public on the Global Voices website.

After initial byline analysis using OGT, a spreadsheet of identified and unidentified names was shared with Global Voices editors for additional gender identification, with instructions to flag accounts whose gender should not be tagged or anonymity was required. In conversation with Global Voices, data was cleaned automatically to include the correct names in cases where usernames and translator names had been substituted for author names. Posts were segmented into two groups based on length. Global Voices has two main kinds of posts: short paragraphs linking to other news sources and more substantial blog posts. Link posts, which are mostly authored by paid editors, were omitted from this analysis.

Compared to other publications, Global Voices shows an unusual gender diversity for a news site. After initial fluctuation in its first few years, Global Voices has settled into a consistent ratio of content by men and women. From 2009 to 2012, the percentage of posts by men and women stayed within 6% of each other.

![Figure 24: Author gender in Global Voices, 2005-2012](image)

Across Global Voices, among contributors who have published many posts to contributors with only a few posts, far more women are contributing posts than in
mainstream media. Within these brackets, men still outnumber women at all points, sometimes by as much as eleven percent. 52% of contributors who publish 9-23 posts are identifiably male, while 41% of women in the same bracket are identifiably women. Across sections analyzed, women outnumbered men, except in the Sub-Saharan Africa section, where women still wrote 40.8% of posts.

Figure 25: Author gender in Global Voices at different levels of participation

Figure 26: Author gender in Global Voices across sections
In an interview, Solana Larsen, the editor of Global Voices, suggested that although Global Voices never set gender parity as an explicit goal, increased parity in terms of international representation is an explicit goal of the project. This worldview and culture likely contributed to the diversity among its contributors. Global Voices, she said, attracts contributors who "see the world in terms of shared experiences and similarities rather than differences." It's common for Global Voices contributors to write about gender politics even if they're not the subject of a story. Men write about feminism and straight people write about LGBTQ stories. Gender balance among editors and leadership does play a role, but perhaps more important is the cultural focus on collaboration and non-competitiveness within the team. Another possible reason for the diversity in Global Voices is its focus on curation rather than opinion. Global Voices editors are chosen for their interest in highlighting and sharing other people's work. Solana also identified an emerging trend for Global Voices translators to start writing original posts, which might influence diversity on the site in the future.

In this case study of Global Voices gender diversity facilitated by Open Gender Tracker, we were able to show that women have contributed at all levels of participation, across studied regions and time. The data also fostered a productive conversation about patterns of success within Global Voices. Within a citizen media site like Global Voices, this conversation can be helpfully informative for editors and contributors alike.

In the future, OGT could be used to investigate conversion rates from translators to full contributors. With a new parser to load RSS feeds, it could also keep editors informed with up-to-date information about the diversity of contributions. Perhaps a live system might offer more constructive feedback than was supplied by this study. Finally, lessons from the Global Voices study could be used to create a "success detector" with OGT that searches the archives of thousands of publications in search similar case studies to inform a comparative, qualitative exploration of the factors supporting gender diversity in citizen media.
7.3.2 Processing live data feeds: Boston Globe

The Boston Globe, an internationally respected newspaper and the premier regional newspaper of New England, participated in a case study focused on integrating Open Gender Tracker with their newsroom's software. The Globe makes excerpts of its content available via a machine-readable API developed by Chris Marstall at Globe Labs. In this case study, web-based software was developed to apply OGT to interactive queries by users. This interactive querying and visualisation software was able to support filtered exploration and discussion by people internal to the Globe. In the future, this kind of interactive exploration may be used to explore and prototype newsroom gender reporting systems.

To operate the Boston Globe API gender tracker API explorer, users operate the main Boston Globe API page to find a query for which they want gender data, specifying facets such as author, section, page, topic, and date. Next, users paste that query URL into the gender tracker client, which fetches the API results, parses them into Open Gender Tracker, attaches gender data to the result data, and stores the data in a results cache. Results are shared to an interface that presents a summary of article gender, a timeseries chart of content gender, a classifier of author gender, a list of gender judgments per post, and links to raw data for each query that documents the probabilities leading to a particular gender judgment. Cached queries can be clicked for later exploration.
In a large news organisation with multiple organisational levels and varied metrics reporting approaches within those levels, there can be no single approach to feedback. Flexible querying through APIs offers an opportunity to explore gender data together with those varied parties. Using those reports, it may be possible to imagine tailored gender feedback on the content published by newsrooms.

Open Gender Tracker Interactive API Explorer

( montage of interface elements, articles mentioning Beer Jan - March 2013 )

7.3.3 Case study outcomes
These two Open Gender Tracker case studies represent two technical advances in the automated study of content gender. The Global Voices case study demonstrates the possibility for historical analysis of a publication's entire archives. The Boston Globe
interactive visualization shows the possibility of carrying out live queries on a publication's ongoing content in a way that can be integrated with a publisher's existing systems with little or no modification to those systems.

These case studies also demonstrate the lines of enquiry simplified and automated by OGT. The Global Voices study automates approaches for investigating the diversity of participation in citizen media websites. The Boston Globe API gender integration offers the possibility for users to explore a flexible range of questions within different parts of a news operation.

7.4 Future directions

In the future, Open Gender Tracker could be used to evaluate the ability of content publishers to reach diverse audiences with diverse content. Data on content gender could be paired with data on the gender of Twitter and Facebook accounts that are sharing that content. Perhaps a diversity score could be created to express the degree to which a writer or section is reaching a diverse audience. Future versions of OGT could also incorporate metrics to classify the gender of people mentioned and quoted in the news. Another possibility is the classification of adjectives used to describe women in news content (Brook et al).

When used with APIs like the Boston Globe API, OGT could be used to prototype and build self-monitoring technologies for journalists, a report card which could inform individual journalists about the language they use to describe women, as well as their content's social popularity with men and women. Self-monitoring could also be carried out within sections or across an entire publisher.

When combined with media monitoring platforms like MediaCloud, Open Gender Tracker could be used to search for patterns of Success. MediaCloud accesses the live content of nearly 50,000 media sources, far more than any team could monitor. OGT could be configured to search for publications with a good diversity record, creating a shortlist for further enquiry into what makes diverse newsrooms succeed.
If expanded to include RSS feeds, OGT can be used to support media monitoring efforts. Groups like VIDA Web Women in Literary Arts, the Global Media Monitoring Project, and the Op Ed Project conduct media monitoring through human coding, which is a limiting and laborious process. Automated techniques cannot replicate the nuanced coding of "Who Makes The News" on language, photographs, and the professions of experts. Open Gender Tracker could however completely replicate byline counts currently conducted by VIDA and the Op Ed Project.

The impact of Open Gender Tracker will not be direct. It is a platform for constructing research and transformative technologies which incorporate media monitoring, offering simple interfaces for expanding inputs, metrics, and outputs. Since diversity and inequality vary widely across a great number of incompatible platforms and communities, OGT offers the possibility of tailoring enquiry and theories of change to a given context. These case studies offer two possible directions. Together with Irene and Adam, I hope that this open source software is used for many other impactful purposes.
8. FOLLOWBIAS, AN APP FOR PERSONAL BEHAVIOR CHANGE

FollowBias offers an intervention on women's representation in the media that can be evaluated with a randomized controlled trial. It is a web app that reveals to users the gender ratio of who they follow on Twitter. It is also an experiment to see if making that ratio visible to users can influence who they follow. Users can check their FollowBias over time to monitor change. This record over time can also be used to evaluate the effectiveness of the FollowBias app itself in changing the ratio of who users choose to follow. FollowBias was created in collaboration with Sarah Szalavitz, with graphic design support by James Home.

8.1 Design Context and Goals

When media consumers choose which sites and social media accounts to follow, those choices define the personal biases of the content most readily available to them. Readers who access the news primarily through social media aren't directly affected by what a newspaper chooses to put on the front page of the print or web editions. They are much more likely to see comments, opinions, and links posted by their contacts on social media. FollowBias offers individuals visibility on the voices they choose to see on Twitter.

8.1.1 Social media curation

Many journalists and content curators use Twitter as a primary resource for links to share and voices to amplify. Others curate a conversation on social media to discuss possible story ideas. When these writers and curators choose who to follow, they are constructing the focuses and biases that influence what they amplify and publish to their own audience. FollowBias offers these content producers visibility on those biases.

8.1.2 Randomized controlled trials on social behavior change

FollowBias participates in a growing trend to run large-scale randomized controlled trials to test social interventions. During the 2012 US presidential election, campaigning organisation MoveOn sent voters postcards comparing their personal voting record to their friends, testing the effectiveness of social cues on voter turnout. In a study of voter
participation, Facebook showed some users the profiles of their friends who reported voting and measured their voting participation in contrast with users who didn't see information about which of their friends voted (Bond et al). Facebook has also conducted similar experiments with organ donation, showing users information about others who sign up for organ donation programs (Matias, “Data”). FollowBias, which tracks its users' responses, is designed to support similar experiments on its own effectiveness.

8.1.3 Privacy in bias tracking and social behavior change
FollowBias is also an exploration of privacy in the calculation of potentially-sensitive metrics on public data. The list of who a Twitter user follows is public data, a "bias" score based on that list may be uncomfortable for some people. One goal of this research is to learn more about user preferences and concerns about such numbers and the possibility that those numbers may be seen by others. For this reason, the first version of FollowBias keeps each user's score private to them while collecting feedback on privacy questions.

8.2 Design
The following section describes the design of FollowBias, including its user experience, system architecture, visual design, and a mechanism that coordinates users to improve the accuracy of gender estimates.

8.2.1 User Experience
Users of FollowBias log into the service using Twitter's OAuth authentication. The FollowBias app redirects users to the Twitter site to verify their identity and grant FollowBias access to their public data. If users are part of the study, they are directed to a survey. After completing the survey, those users see the gender ratio of who they follow on Twitter spread into three categories: women, men, or brands, bots, and more. Users are then prompted to take a followup survey and to cooperate to correct the gender judgments of the algorithm. In this final form participation, users review the system's current gender estimates for each account they follow on Twitter. By selecting alternative
gender options, they can correct that estimate, improving the accuracy of everyone's FollowBias ratio.

**Figure 28: FollowBias user experience**

**8.2.2 System architecture**

The FollowBias architecture spreads work across four layers. The browser layer, built in Javascript, manages the user interface, draws the vector graphics display, and performs the basic equations that produce a user's FollowBias ratio. The Controller, a Ruby on Rails application, manages permissions, aggregates and serves data used by the browser layer, determines which users are allocated to the control or treatment groups, and manages the system for user corrections to gender judgments. The Job Queue schedules and executes background processes to query the Twitter API for the public list of which accounts a user follows and assign gender scores to any accounts that it hasn't yet seen.
8.2.3 Visual design: gender binaries and 3D glasses

By presenting a user's gender ratio in the form of 3D glasses, FollowBias offers an argument for parity while also foregrounding the constructed nature of the information it collects and shares with users. To see the world in full perspective, one needs to see through more than one standpoint. While 3D glasses implicitly encourages diversity, it also calls into question the metrics within its lenses. The feedback offered by FollowBias, like filtered lenses, is a work of artifice and presentation, an imperfect filter between users and the media through which they interpret online media.
A user's FollowBias score is presented as a pie chart of women, men, and a third category, brands bots and more. Accounts with first names that can be classified using demographic name data are classified as men or women. A large number of accounts do not include a full name in Twitter's "name" field. Many of those accounts, such as "The New York Times" and "horse_ebooks" are brands and bots. Many people whose Twitter account predates the availability of the "name" field have not associated a full name with their account. Other accounts omit name information to maintain anonymity. All of these accounts are labeled "Brands, Bots, and more." Non-binary genders are included in "brands, bots, and more" to prevent sensitive gender information from being made public through the classification system.

8.2.4 Correcting FollowBias accuracy

Users are asked to review and correct the accuracy of their FollowBias score. An introduction to the corrections interface explains the classification system and asks users not to reveal gender information that users have not disclosed. Before starting, users are shown examples of classified accounts from the list of who they follow. When users choose to correct gender estimates they are shown a long-scrolling list of every Twitter account they follow. Accounts with likely incorrect estimates are sorted to the top of this list. To correct the gender estimate, a user clicks on an alternative classification, which is stored to the server. That user's FollowBias is automatically updated on screen every time a correction is made. At present, an account's gender is taken from the last user correction associated with that account.
Users are only granted access to their own FollowBias score. The service shows them their latest FollowBias score as it changes over time. After a user signs up for FollowBias, the software also checks Twitter every six hours for information on any changes to who a user follows on Twitter. For each user, the software keeps records from the moment that the user was added to the system, the moment that the user first saw the FollowBias score, the end of that user's corrections process, and any subsequent changes in who that user follows. At each of these moments, the reporting software can calculate the score visible to the user as well as the closest current estimate of the absolute scores, based on subsequent corrections to account genders.

8.3 Participant Responses
In an initial study with 63 users, we explored four areas of inquiry. Would people trust their FollowBias score and participate in correcting account genders? How do users interpret their FollowBias score in relation to their social context and use of Twitter? Would users change who they follow upon being exposed to their FollowBias? Finally, we collected feedback on privacy concerns for users of FollowBias.

8.3.1 Study design
Study participants were recruited from a list of journalists, bloggers, and other active Twitter users who publish original content or regularly retweet content from others. Participants were sent an email inviting them to click a link and try FollowBias. After logging into FollowBias using Twitter, all participants were shown a survey. Upon
completing the survey, the treatment group was shown the FollowBias app. The control group (19% of participants) were shown a picture of Dubstep Cat wearing 3D glasses and exposed to similar text about Twitter bias as the treatment group. The treatment group, but not the control group, are given a followup survey.

8.3.2 Study participants
Among all participants, 42% were male and 55% identified as female. 50% of participants self identified as journalists, and 45% considered themselves bloggers. Another 30% considered themselves academics. Most participants use Twitter hourly, with 94% using Twitter at least daily. Mobile apps are the primary way for 53% of participants to access Twitter, with another 44% accessing it primarily from the desktop. More than half of participants regularly create and share original content, curate and share content from other sites, engage in conversations, find sources for content they create, and consume content via links. 42% of respondents reported using some kind of tool to track social media metrics, and 13% report tracking the gender of their Twitter audience. 10% already report tracking the gender of who they follow on Twitter.

8.3.3 Trusting FollowBias, making corrections
Did users trust their FollowBias report? In followup email and the followup survey, most participants reflected on the personal and social reasons for their score rather than questioning the metric. Those who disagreed with the score would have preferred a category for pseudonyms or non-binary genders. Several participants emailed us to let us know that they felt that too many people were classified as brands, bots, and more, and that they appreciated the ability to make corrections.

The corrections interface was used by 23% of participants, who made a total of 3,254 corrections. The greatest number of corrections were made by a participant who follows a large number of non-US accounts. That user made 337 corrections, 78% of the entire set of people that user followed at the time. Most users corrected less than 20% of the list of who they follow. Users tended to carry out corrections in a single session taking between
5 and 25 minutes. Five users carried out their corrections over a period longer than an hour, perhaps in a casually stretched-out way.

8.3.3 Interpreting a FollowBias score

When asked to explain the complexities that influenced the outcome of their FollowBias score, participants sometimes cited professional pressures versus their preferred personal behavior. "Who I follow, in part, is a function of my professional networks," wrote one participant, citing a male-dominated professional environment. One participant emailed us screenshots of the difference between a personal and professional account. This participant's professional account followed 36% women and 16% brands, bots and more, while the same participant's personal account followed 54% women and only 1% brands, bots, and more.
That participant writes,

The performance of people I need to follow for politeness and algorithmic stuff is more male. The people that I actually follow and read every day is more female. [....] Key to this is acknowledging that I work in the tech sector and that there's a lot of politeness involved with following people so as to not offend. And since the tech sector is predominantly male, this bias is visible.

This distinction between personal and professional Twitter accounts was common for many participants. In the opening survey, 40% of users reported distinctions between professional and personal Twitter uses, although only a quarter of those kept both a professional and personal account.

Another respondent also felt pressured by metrics systems to follow accounts that might not be actual interests: "the problem with my followers on this account above all else is that it's a performance for other algorithmic analyses, not actually indicative of who I pay attention to." The particular metrics system isn't listed, but "followback" software is be a typical example of software which helps social media users choose who to follow. Such metrics often try to maximising a Twitter user's audience and the reach of that user's content.
8.3.4 FollowBias privacy concerns

Several participants shared privacy concerns. One journalist employed by a mainstream news organisation wrote, "I'm slightly nervous. The organisation I work for prides itself on being objective and I take that value seriously in my work." Acknowledging that FollowBias uses entirely public information, the participant remarked that "if you're making the process really easy and calling it "FollowBias", it might be a bit uncomfortable if that then got published with my name attached to it." Another participant, who runs a prominent feminist Twitter account, also requested that the FollowBias score be kept private. "We already receive a lot of abuse from men," wrote this participant in an email, "revealing the demographic of who we follow on Twitter (probably mostly women) would be likely to increase that and open us up to further criticism and accusations!"

8.4 Measuring change

Although experiments and data analysis are ongoing, some preliminary data analysis is available. Follow data was collected for each participant before recruitment using public information on who they follow on Twitter. The FollowBias reporting system can report historical data using the most up-to-date set of gender judgments based on user corrections. It can also recreate the percentages seen by participants after completing the survey and after completing all of their corrections. This perception data is not analyzed here. Start and end dates for the study are also inappropriate to include here, since participants chose to log in on different days and the dataset included here is relative to each participant's experience.

Before participating in FollowBias, half of FollowBias participants followed at least 20% more men than women. Only 16 participants followed more women than men.
Figure 35: Difference between women % and men % across all participants

Data on the social graph and FollowBias score for each participant was recorded every six hours starting before participants were recruited. The following calculations and charts of changes in FollowBias score compare the score before participation to data collected several weeks later. The score used is the change in the difference between the percentage of women followed and the percentage of men followed. A positive number indicates a change towards a greater percentage of women. A negative number indicates a change towards a lesser percentage of women. This is an imperfect score for a study on diversity, since it only measures changes towards a greater proportion of women. Nevertheless, this score does illustrate the kind of analysis possible with FollowBias.

Among the control group, 46% of participants increased the percentage of women they follow after exposure to the survey about gender in social media. Among the treatment group, only 42% of participants increased the percentage of women they follow after exposure to the survey and their FollowBias score. The following charts show changes in the difference between women and men follows for the treatment and control groups before and after exposure to the web app.
This initial analysis does not prove or disprove the hypothesis that exposing participants to their FollowBias score prompts users to follow more diverse genders on Twitter. The score does not foreground a move towards diversity, nor does it take into account the full depth of timeseries data available within FollowBias.
8.5 Future Directions

A more effective FollowBias will be able to track a greater range of Twitter behavior. Retweets amplify voices and could be measured in terms of how far they spread. Tweets could be analyzed for the diversity of the audience that tweets and retweets what a user posts. Using Open Gender Tracker, the byline and subject gender of shared links could also be tracked and made visible to FollowBias users.

It is likely that the most effective system will make recommendations and offer a simple interface for users to change the diversity of the voices they follow on Twitter. Recommendation systems may be editorially produced, crowdsourced through a hashtag or voting system, assembled from Twitter's own recommendations, or drawn from a user's social graph. Further experiments can evaluate the effectiveness of these recommendation approaches.

Establishing causality for FollowBias will be difficult to establish without the ability to access recommendations made elsewhere. Twitter users can be influenced by the list of who follows them, who mentions them, who is mentioned by their friends, who is mentioned in Twitter's automated emails, and who is suggested on Twitter's website.

Another set of features could show users the FollowBias and the progress of friends who are taking steps to change their FollowBias score. Studies like the Facebook Voter Participation study suggest that selectively showing users the behavior of peers can be an effective way to influence users' choices. Other possibilities include allowing users to compare themselves to each other and challenge each other. All of these possibilities involve sensitive privacy decisions to avoid conditions favorable to bullying and hate speech.

FollowBias is a showcase of an intervention for women's representation that can be evaluated with a randomized controlled trial. In addition to raising awareness about the role that everyone's social media activity plays in the representation of women, it is a platform for attempting and evaluating approaches to increase the diversity of voices that
people read and amplify online. This first experiment establishes a platform for many further features and experiments.
9. PASSING ON, USING DATA FOR PARTICIPATORY PARITY

Passing On, designed with Sophie Diehl, is an interactive data visualization that coordinates viewers to use their own voices to discover and address the disparity of women's visibility in the media. After showing viewers a visualization that makes clear that women are a small minority of the people who have appeared in New York Times editorial obituaries, Passing On encourages viewers to check for gaps in Wikipedia that can be filled with those obituaries. The software guides viewers from small, meaningful actions like reading an obituary or sharing links, to more substantial actions like identifying gaps in Wikipedia, requesting new articles, or even creating articles for women who don't yet appear in Wikipedia. All of this activity is shared in a cooperative effort across all viewers of the site.

![Figure 38: Passing On, by J. Nathan Matias and Sophie Diehl](image)
9.1 Disparities in life records

Passing On bridges between disparities that are hard to change through collective action and disparities that are directly addressable by collective action. Content about women is a minority of what appears in both the New York Times and Wikipedia. Few of us can do anything to directly change the proportion of women featured in the New York Times. Anyone can, in principle, add more women into Wikipedia. Passing On uses data from the New York Times to acknowledge disparities and support change that we can create together.

9.1.1 Gender disparities in New York Times obituaries

Women remained a small minority in the New York Times obituary section over decades. Between 1987 and 2007, obituaries about women were consistently less than than 20% of obituaries in the New York Times, with the majority of those paid death notices. In the subset of 35801 obituaries included in Passing On, only 5674 are about women, 15.8% of the total set.

9.1.2 Gender disparities in Wikipedia

The English language edition of Wikipedia shows a gender bias that is similar to the New York Times, with articles about women comprising 16% of biographies. This is to be expected, since someone needs to appear multiple times in the mainstream media to be eligible for inclusion in Wikipedia. English Wikipedia does have biographical entries for more women than any other encyclopedia. Nevertheless, when compared to other datasets of notable women, Wikipedia is less balanced in who it neglects to include than Britannica. The following diagram illustrates the kind of difference that exists between Britannica and Wikipedia, in relation to the set of possible biographical identified in research by Joseph Reagle and Lauren Rhue:
9.1.3 Parity beyond balance

A narrow focus on gender ratios can treat visibility as a limited resource, as if more visibility on one side necessarily results in less visibility for the other. That may be true in cases of employment or print media, where the cost of hiring employees and the cost of printing create hard limits on the space of opportunity. Perhaps for the New York Times, featuring more women would involve publishing about fewer men. Sites like Wikipedia are not constrained by publication size or the cost of hiring writers. Instead, women's visibility in Wikipedia is constrained by the pool of active contributors interested in adding women to Wikipedia and the availability of records that establish the notability of potential biography subjects. Passing On uses the New York Times dataset to expand both resources. When Passing On users cooperate to identify people in the New York Times to add to Wikipedia, they learn more about participating in Wikipedia and take first steps to contribute.
9.2 Design history

Early designs of Passing On began as an exploratory data visualization by Sophie Diehl entitled Gender in Memoriam. Sophie used the PUGG system to group New York Times editorial obituaries into profession, relationship, and life story categories. Gender in Memoriam directly compares men's and women's obituaries, showing comparative volume over time and inviting users to read sentences from those obituaries. Passing On is a further elaboration on Sophie's data analysis, extending it with a theory of change focused on adding women to Wikipedia and helping readers to learn how to participate in Wikipedia.

Figure 40: Gender in Memoriam, by Sophie Diehl
9.3 User Experience

"Passing On" starts with a slideshow that asks viewers to reflect on the heroes that inspire our aspirations and the possibility that women's limited visibility influences who we choose as heroes. The slideshow also encourages readers to explore New York Times life records, share stories, and update Wikipedia to "write and remember women's lives, together." Next, viewers are shown a visualization of obituary gender in the New York Times for 49 different topics ranging from "business" to "sport" and "awards". When users click on a search, Passing On shows them a graphical representation of all women's obituaries returned by the query, color-coded by how much activity other viewers have taken to classify and add those people to Wikipedia. When users select a person, they are presented with basic information about that person, including some quotations from the New York Times Obituary.

From the point that users select an obituary, further user actions contribute toward a collaborative effort to make women's lives visible. After users follow a link to the New York Times website, they are encouraged to share the link and comments with their friends on Facebook and Twitter. After users look up a person on Wikipedia, they are
prompted to answer if Wikipedia has an entry for that person, if that person meets Wikipedia's requirements for inclusion, and if that person's existing Wikipedia page could be improved. After users look up a person in other publications, such as Harper's, Time Magazine, or the Boston Globe, Passing On records the web location those articles.

All of these acts of searching and browsing add to a combined record that can be used to improve an existing Wikipedia article or start a new one. The status of that combined record is visually displayed back on the the grid of people appearing in New York Times obituaries. As more users participate, the grid fills with color. Every point in the user experience is associated with a specific URL that can be shared with others, either to highlight someone's story or to request help compiling data for a Wikipedia entry.

Cooperation within Passing On ranges from the casual action of reading an obituary to the very involved of authoring a Wikipedia article. Even the most casual action is a meaningful contribution. Upon completing each action, users are invited to try something more substantial. Reading obituaries leads to sharing them on social media. Discovering the absence of a person in Wikipedia leads to information about the process for adding someone to Wikipedia. Finding a Wikipedia article about a person leads to reflection on what makes a good Wikipedia article and questions about the quality of the page. When users discover absent and inadequate pages, users are encouraged to perform searches on other sites to compile the information required to create and update a biographical entry on Wikipedia. Upon learning these skills, users are encouraged to repeat the process for others in the New York Times obituary corpus. Finally, Passing On invites users to use the collected information about a person to create a new Wikipedia article. Users need not carry out the most time-consuming activities to make a substantive contribution; all actions in the system carry the potential to improve Wikipedia's coverage of women.

9.4 System Architecture
The system architecture of Passing On links together four resources with a central web application. A news data processing system provides data to visualizations. A javascript application in the web browser manages visualizations and user interaction. A server-
based activity archive aggregates and reports data on user collaboration. Finally, users access and report on information resources from external websites.

Obituary data is collected by processing the New York Times Annotated Corpus with PUGG, an open source Python-based news data analysis platform designed for this thesis, which was used for UK news analysis, and which was the precursor to Open Gender Tracker. PUGG classifies the gender of each obituary and produces pre-cached search results for the New York Times obituary visualization. Those results are served from a web server to the Passing On browser application, which presents the user interface.

The browser interface, a javascript application, manages the user experience of Passing On. It presents a visualisation of New York Times obituary content gender\(^1\). It directs users through the process of viewing available obituaries to read more about women's lives. Finally, it manages the process of collecting information from users that can be used to fill gaps in Wikipedia.

---

\(^1\) The bubble cloud visualization of obituary gender was inspired by Bostock, Carter, and Ericson's US National Convention speech visualization in the New York Times. Bubble cloud software was adapted from an open source demo created by Jim Vallandingham.
The Passing On web application is an archive of cooperative activity. It accepts data from users and sharing aggregate information to fill in the grid of crowd activity. The web application also collects information about the overall contributions of individual browser sessions, supporting the possibility of showing users their personal actions in context of collective progress.

Passing On also relies on a network of archival and search systems on third party sites which permit instructions by URL query string. It simplifies the process of verifying the eligibility of a person for inclusion in Wikipedia by converting searches across 8 publications into direct links. Follow-up forms collect bibliographic information on a person's appearances in those publications, which is then archived to the Passing On server application.

9.5 Future Directions

Passing On still needs substantial improvements before public launch. An effective campaign will include a mechanism to establish ongoing contact with users. Followup emails and tweets could offer participants feedback on the outcome of their work and invite them to be involved in ongoing participation. Search functionality would help users focus on improving women's visibility in Wikipedia for areas where users have the greatest expertise. A better feedback system and contribution pipeline could channel participation more effectively and provide clearer feedback to all participants on the status of everyone's work and their individual contribution the collective project. The most effective project would require coordination with Wikipedians and other organisations who are committed to incorporating contributions into Wikipedia articles. Finally, data from the web application can be sent in machine readable form to data platforms such as DBPedia and WikiData to improve women's representation in machine readable systems as well.
10. NETWORK TACTICS FOR REPRESENTATION IN THE NEWS

In this thesis, I argue that automated content and network analysis is a powerful approach for understanding and changing women’s representation in the news. I present example metrics with case studies that demonstrate techniques and perspectives of interpretation. Open Gender Tracker, FollowBias, and Passing On represent three applications of those techniques into designs to support and evaluate change towards a media that represents women more fairly. These designs illustrate the value of careful conversation with each context, flexible support for personal interpretation, and cooperative tactics that repurpose environments of disparity in constructive directions. Qualitative results from case studies and user testing illustrate the value of nuanced approaches to privacy. Some critical design questions remain unanswered however, on the definition of fairness and the palette of tactics suggested by this work.

10.1 Defining and measuring fair representation

What constitutes a definition for fair representation in networks, where we expect the voice of women to vary across different parts of the Internet? Without a definition of fairness, measurable change in women’s visibility cannot easily be described as progress. Although I offer no definition of network pluralism, the cases, designs, and results in this thesis are suggestive of the requirements for such a definition.

10.2 Further Design Ideas

Open Gender Tracker, FollowBias, and Passing On are just three examples on a palette of software tactics for women’s representation in the news. Each example focuses on a particular point of intervention in the news: obituaries, APIs, and Twitter. Each example applies a different approach to change. FollowBias is a personal tracker for behavior change. Passing On facilitates cooperation. Open Gender Tracker is a toolkit to support news organisations that want to monitor gender and the advocacy organisations that address the same issues. The following project ideas are offered as inspiration for further technologies designs to support the visibility of underrepresented people:
**Who They Recommend** is a technology that monitors recommendation systems for the diversity of what they suggest, making the biases of those recommendations visible to users. The most basic version follows the Twitter and Facebook accounts of publishers, as well as scraping the parts of websites that rank and arrange content preferentially. A more sophisticated version permits users to log in and monitor personalized recommendations.

**Fairness Sketchpad** is a Wordpress plugin to give journalists feedback on the language they use to describe people in their articles.

**Diversity Check** takes audio, text, or video input from the draft of a news story and offers multiformat newsrooms feedback on the gender diversity of their content in time to adjust show contents.

**diversity.reading.am** is a module for personal reading trackers like Reading.am, that tells you the gender diversity of content read by you and your friends and points you to curators whose reading habits are closer to your personal goals.

**CareerTracker** is a technology for tracking an organization’s success at supporting the careers of writers over time. It collects data on what those writers publish elsewhere and aggregates reports on the rate of success an organisation has in placing writers in writing opportunities.

** Literary Gender Tracker** is an automated system to monitor gender in book reviews, book clubs, promotions, and sales across the book trade.

**Daily Diversity** is a news recommendation system that invites readers to select the kind of speaker diversity they seek and include content on a wide ranging set of topics, pulled from across the web, that matches the reader’s diversity goal.
**Public Faces** is an app for monitoring the diversity of people speaking in press releases and marketing material. As a business, it could help companies test the connection between diverse marketing and diverse customers.

**diversity.lanyrd.com** is a monitor for event management systems to track the diversity of attendees and invited speakers at conferences and unconferences.

**diversity.thanks.fm** is a module for acknowledgment trackers like thanks.fm to inform creative collaborators about the diversity of who they work with and how they choose to acknowledge collaborators publicly.
REFERENCES


American Society of Newspaper Editors Newsroom Census, 1997-2012  


<http://jmq.sagepub.com/content/81/1/139.short> Accessed March 21, 2013


Barnett, Rory. "Politicians woo 'Mumsnet' generation' Feb 18, 2010  


Belam, Martin. "How do British newspapers compare to Newsweek's catastrophic 51% circulation collapse?" Oct 22, 2012  


<http://www.nature.com/nature/journal/v489/n7415/full/nature11421.html>

Bostock, Mike, Shan Carter, Matthew Ericson, "At the National Conventions, the Words They Used." New York Times, Sept 6, 2012.  

BlogHer, "Women and Social Media in 2012"  

102


Larsen, Solana. Interview March 5, 2013, with Irene Ros and Adam Hyland.


Linguistic Data Consortium (NY Times Data) Accessed June 3 2012


Ros, Irene. J. Nathan Matias, Adam Hyland, Ami Sedghi, "GenderTracker: Global Voices Gender Balance Case Study” <https://docs.google.com/spreadsheet/ccc?key=0Al5UYaVoRpW3dHBRVEIzIXzIDV3FH0VN3U1NfQVRkTmc#gid=0> Accessed 11 May 2013


Stierch, Sarah. "We’ve already had numerous articles nominated for deletion. But we’ve saved them. But I mean really..? I swear... #women #wikipedia" 30 Mar 2012 <https://twitter.com/Sarah_Stierch/status/185819624015798272> Accessed March 25, 2013

