White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/110764/

Version: Accepted Version

Article:
Ford, H and Wajcman, J (2017) 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. Social Studies of Science, 47 (4). pp. 511-527. ISSN: 0306-3127

https://doi.org/10.1177/0306312717692172

© The Authors 2017. This is an author produced version of a paper published in Social Studies of Science. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

‘Anyone can edit’, not everyone does: Wikipedia’s infrastructure and the gender gap

Heather Ford
School of Media and Communication, University of Leeds, UK

Judy Wajcman
Department of Sociology, London School of Economics, UK

Abstract
Feminist STS has long established that science’s provenance as a male domain continues to define what counts as knowledge and expertise. Wikipedia, arguably one of the most powerful sources of information today, was initially lauded as providing the opportunity to rebuild knowledge institutions by providing greater representation of multiple groups. However, less than ten percent of Wikipedia editors are women. At one level, this imbalance in contributions and therefore content is yet another case of the masculine culture of technoscience. This is an important argument and, in this article, we examine the empirical research that highlights these issues. Our main objective, however, is to extend current accounts by demonstrating that Wikipedia’s infrastructure introduces new and less visible sources of gender disparity. In sum, our aim here is to present a consolidated analysis of the gendering of Wikipedia.

Keywords
gender, infrastructure, platforms, expertise, technoscience, feminism, Wikipedia

Correspondence:
Heather Ford, University of Leeds – School of Media and Communication, Clothworkers North, University of Leeds, Leeds, West Yorkshire LS2 9J5, UK
Email: h.ford@leeds.ac.uk

Introduction

Wikipedia has a gender problem. While exact numbers are difficult to estimate, no one disputes that the overwhelming majority of contributors are male (Glott et al., 2010; Hill and Shaw, 2013; Wikimedia Foundation, 2011). That this is an intractable problem was recently acknowledged by Jimmy Wales (2014), co-founder of Wikipedia, who admitted that the Wikipedia Foundation had ‘completely failed’ to meet its goal of increasing the number of female participants to 25% by 2015. Social media sites such as Facebook and Twitter do not exhibit this gender disparity, raising questions about Wikipedia’s recruitment and retention practices. The resulting gender bias in content coverage is also increasingly recognized. Wikipedia’s gender gap has thus emerged as a topic of interest and concern for scholars of new media and related fields.

The question about the systemic bias of the encyclopedia is important because Wikipedia has become one of the most powerful global media of our time. Read by 365 million people around the world and available in 275 languages, Wikipedia is the sixth-biggest website in the world. And despite its dwindling (or stabilizing – depending on your interpretation of the figures) *editor* base, its *readership* is steadily increasing, as many developing countries come online and are provided with free access on mobile phones through the Wikipedia Zero program (Koetsier, 2013). Furthermore, Google and other search engines' prioritization of Wikipedia content in search results and in 'fact boxes' means that facts produced within Wikipedia are increasingly fading into the background and becoming black-boxed, so that readers often accept Wikipedia's representation of the world as natural and obvious.

Yet this is an online encyclopedia that prides itself on being open and collective, promising to democratize knowledge institutions by enabling amateurs to participate in the representation of knowledge. The Wikipedia logo is of a spherical puzzle, still unfinished. The project is thus constituted as an objective knowledge project that must merely be filled in. Wikipedia's goal is to work towards 'the sum of all human knowledge' and its rhetoric focuses on the gaps in knowledge still to be filled. How can it then be that Wikipedia, the former poster child of the internet, has a gender problem?

While some authors frame the issue in terms of women's lack of technical skills and confidence, the adversarial culture of the Wikipedia community has itself become a focus of attention among researchers. Gender studies of Wikipedia have thus made some important contributions to understanding this inequity, as we will elaborate below. Yet, to date the discussion has not been informed by the rich tradition of feminist STS that substantively foregrounds parallel arguments. An initial aim of this article, then, is to provide a broader conceptual grounding for contemporary analyses of the gender gap in Wikipedia.

Our main objective, however, is to extend these accounts by demonstrating that Wikipedia's infrastructure introduces new and less visible sources of gender disparity. Specifically, we argue that Wikipedia's origins and the infrastructures on which it relies are based on foundational epistemologies that exclude women, in addition to other groups of knowers whose knowledge does not accord with the standards and models established through this infrastructure.[1] Wikipedia is built on an installed base (Star, 1999) of existing infrastructures – in this case, infrastructures from the modern encyclopedic project and the open Internet. While the encyclopedia has inherited some biases from those projects, the relations that define Wikipedia have grown incrementally over time in a reconfiguration that produces bias that is unique to the project. By drawing on a long-term ethnographic study of Wikipedia (Ford, 2015) and feminist STS, we demonstrate how Wikipedia's infrastructure produces hidden layers of gendering at the levels of code, policy and logics that can only be solved by greater diversity in those configuring such infrastructure. In sum, our intention here is to present a consolidated analysis of the multiple levels of the gendering of Wikipedia.

Background: From the deficit model to gendered technoscience

Charting the recent debate about the paucity of female Wikipedians (that is, editors), it is striking how familiar the diagnosis and solutions are. So it is worthwhile to recall some of the earlier discussions, especially as gender studies of Wikipedia appear to be developing along similar lines.

Since the 1970s, feminist scholars have been documenting and explaining women's low participation rates in STEM (science, technology, engineering and math) subjects and professions (Long and Frank Fox, 1995; Rossiter, 1982; Rothschild, 1983). Many studies identified structural barriers to women's participation, looking at sex discrimination in employment and the kind of socialization and education that girls receive which have channeled them away from studying mathematics and science. Explaining the under-representation of women in science education, laboratories and scientific publications, research highlighted the construction and character of femininity encouraged by our culture.

Much early second wave feminism posed the solution in terms of getting more women to enter science and technology – seeing the issue as one of equal access to education and employment. Rather than questioning technoscience itself, it was generally assumed that science was intrinsically open, concerned with unbiased and objective research. If girls were given the right opportunities and encouragement they could easily become scientists and engineers. Remedying the gender deficit was seen as a problem that could be overcome by a combination of different socialization processes and equal opportunity policies.

This liberal feminist tradition located the problem in women's aspirations and values, rather than asking broader questions about whether STEM disciplines and institutions could be reshaped to accommodate women (Wajcman, 1991). The equal opportunity recommendations, moreover, asked women to exchange major aspects of their gender identity for a masculine version without prescribing a similar 'degendering' process for men. In order to succeed, women had to model themselves on men.

The limited success of these equal opportunity strategies was a catalyst for a more critical feminist STS that saw women's reluctance 'to enter' as to do with the sex-stereotyped association of science and technology as activities appropriate for men. It was not simply a question of acquiring skills, because these skills and competencies were embedded in a culture of masculinity that was largely coterminous with the culture of technoscience (McNeil 1987). Over the next decades, the feminist critique of science would evolve, as Harding (1986) famously expressed it, from asking the 'woman question' in science to asking the more radical 'science question' in feminism, and technical knowledge and expertise would similarly be critiqued as distinctly masculine projects (Balsamo, 1995; Berg and Lie 1995; Haraway, 1997; Rose, 1983; Wajcman, 1991, 2004).

Consequently, an early project of second wave feminism was to recover women scientists and inventors 'hidden from history'. The publication of biographical studies of great women scientists, such as Rosalind Franklin (Sayre, 1975) and Barbara McClintock (Keller, 1983), served as a useful corrective to mainstream histories of science, demonstrating that women had in fact made important contributions to scientific endeavour. Histories of women inventors soon followed, and the legacy of engineering was scrutinized as foundational to the contemporary male ethos of engineering and computing (Cockburn, 1985; Faulkner, 2001; Henwood, 2000). Indeed, as Cowan (1979) vividly revealed, the very definition of technology was cast in terms of male activities: hence the plentiful histories of the steam engine as opposed to the baby bottle. Recovering the history of women's achievements became an integral part of feminist scholarship in a wide range of disciplines. While this literature would subsequently be criticized for its tendency to treat masculinity and femininity as binary, essentialist, stable, heteronormative categories (e.g., Landstrom, 2007), it did establish an approach that conceived of gender and technology as mutually constituted. In other words, it moved the analysis away from treating people and things as distinct entities to the idiom of technology as a sociomaterial or sociotechnical practice.

Within the Wikipedia community, the likelihood of imbalanced topical coverage is also increasingly being acknowledged. Over a decade ago, the Wikiproject 'Countering Systematic Bias' pointed to how the interests and the demographics of its contributors affected its topical coverage (Wikipedia, 2004). The popular press has recently featured stories about how male-linked subjects, like video games and porn stars, have much more coverage than traditionally female-linked interests. In *New Statesman*, for example, Kleeman (2015:35) compared the 'List of Pornographic Actresses' to the 'List of Female Poets'. While the former is 'meticulously referenced, with clear sections according to decade', the latter is 'a sprawling dumping ground' and, while '(t)he list of poets has been edited 600 times, by nearly 300 editors, the list of female porn stars is a newer page but over 1,000 editors have edited it more than 2,500 times.'

At one level, this bias is surprising as, compared to a conventional encyclopedia, 'anyone can edit'. The obvious comparator is the conventionally produced Encyclopedia Britannica, whose 11th edition had very few women contributors (2%) and no women listed among the 49 editorial advisors (Thomas, 1992:18). How then does Wikipedia's coverage of women's biographies, for example, compare to the iconic Britannica? It turns out that Wikipedia does have significantly greater coverage than Britannica. In absolute terms, Wikipedia has more biographies of women. However, the authors of this study (Reagle and Rhue, 2011) conclude that there still is gender bias. This is because Wikipedia's missing articles are disproportionally about females relative to those of Britannica, as evidenced in the ratio of female to male biographies in each work. Given Wikipedia's position as a central reference source, the invisibility of these missing entries means that women are being written out of history once again.

This focus on the 'missing women' in Wikipedia has led to various remedial actions.

Feminist hackathons aimed at improving coverage of women in Wikipedia in the arenas of science, technology, history and the arts have taken place in cities around the world. The British Royal Society conducts annual editathons surrounding Ada Lovelace Day, local WikiWomen groups have regular meetups and Women's History Month is celebrated by Wikipedians in order to redress skewed coverage. These actions directly mirror the early attempts by second wave feminism to restore women to their rightful place in the genealogy of technoscience.

Unsurprisingly, many studies trace the origins and causes of the low numbers to the wider gender imbalance in computer-related fields. The research literature suggests that this disparity is largely due to women's lack of skills, confidence, fear of criticism and conflict (Cassell, 2011; Herring, 2003). On the basis of a large cross-national user survey of Wikipedia, Collier and Bear (2012: 390) find strong evidence that men are more confident than women about their knowledge and expertise, and that women were put off 'due to the high levels of conflict involved in the editing, debating, and defending process'. Synthesizing much of the research, the authors argue that, relative to men, women avoid conflict, competition and negotiation, preferring to share and collaborate (see also Iosub et al., 2014). Although the authors advocate a change in Wikipedia's confrontational style, their psychological approach effectively frames the issue as a deficiency in women themselves. Even Hargittai and Shaw (2015), who stress that the gender gap in editing is only found at high internet skill levels, leave us pondering as to why it is that these women underrate their skills and so are much less likely to contribute to Wikipedia than are comparably skilled men. As this is one of the few articles that documents differences between women, we could not pursue the issue of whether and if so, how, women's underrepresentation varies by, for example, class, sexuality, race and ethnicity.

Ironically, Wikipedia's idealized freedom rhetoric is, in practice, used to dismiss concerns about the gender gap as being merely matters of preference and choice. Tkacz (2012), for example, argues that discourses relating to collaboration and openness on Wikipedia tend to be depoliticized and that there is an invisible politics at work on Wikipedia – the 'politics of the frame' – that necessarily excludes certain points of view. Well-worn debates about the pros and cons of freedom vs. censorship and structurelessness vs. formal systems are rerun within the Wikipedia community.

Recent publications on gender and Wikipedia, however, shift the focus from women to the culture of Wikipedia itself, once more in parallel to former feminist STS discussions. They reveal the extent to which the free, open, libertarian discourse of Wikipedia hides exclusionary practices. As Reagle (2013: 216) points out, 'free culture can be unappealing to those unable or unwilling to hew to the stereotypical features of the online geek (i.e., an identity associated with an intense and narrow interest and argumentative style).' This white, masculine geek identity can be traced back to the computational culture so well-captured in Turkle's writings in the mid-1980s (Turkle, 1984). But Reagle also claims that a new variant has emerged: the 'brogrammer'[2] associated with 'frat-house' culture.

In other words, this contemporary culture is a masculine-only space too, rendering a female geek a contradiction in terms.

No wonder, then, that one of the rare in-depth studies probing the experiences of female Wikipedians found that they must engage in considerable 'emotional labor' in order to participate (Menking and Erickson, 2015). Borrowing Hochschild's (1983) term, the authors report that women strategically avoid certain kinds of work so as not to become targets of trolling, intimidation and harassment by men. Because they value their role as editors, they persist in what they describe as a pernicious environment. Compared to men, women editors have to perform more taxing emotional labor, as being a Wikipedian involves a particularly masculine affective construction of membership. The authors conclude that bridging the gender gap will require more than increasing the number of female editors; it will require Wikipedia to 'better understand its own culture of knowledge production'.

In this article, then, we extend the debate about Wikipedia's gendering from the social characteristics of individual editors to the gendering of Wikipedia's infrastructure. Reflecting developments in feminist theory that interpret gender as ongoingly produced in interaction with technology, we argue that Wikipedia's gendering is constituted not only by the agency of its editors, but also by the deeper logics embedded in the infrastructural relations of the technoscientific project itself. Specifically, we demonstrate how Wikipedia both reinforces and reconfigures gender relations in virtue of a particular assemblage of code, policies and logics of its installed base. Wikipedia's tools and material infrastructure, but most importantly its logics originate from both the free software and Western scientific epistemologies and processes. This has two key implications for gender equity on the platform. First, Wikipedia's identity as an encyclopedia for facts is still governed by historically conservative (male) scientific understandings of expertise and authority. Second, viewed as an infrastructure, Wikipedia requires highly technical expertise, expertise that is traditionally gendered.

Wikipedia as infrastructure

Infrastructure studies and platform studies have been fruitfully mobilized in recent years to improve our understanding of current digital media.[3] In light of significant changes to the ways in which knowledge is digitally produced and mediated, knowledge infrastructures, in particular, have become an important area of study. This move enables Wikipedia to be treated as a socio-technical system, defined as much by its technological tools and managerial dynamics as it is by human actors. To quote Niederer and van Dijck (2010: 1373): 'Wikipedia [is] a gradually evolving sociotechnical system that carefully orchestrates all kinds of human and non-human contributors by implementing managerial hierarchies, protocols and automated editing systems.' As such, Wikipedia not only represents knowledge, but is an active site in which knowledge is produced, setting up disparities in who has access to knowledge sources deemed credible by the Wikipedia's community and in who is accorded the greatest level of authority (Wyatt et al., 2016).

To date, such an infrastructure lens has not been fully applied as a way of understanding the deeper layers of Wikipedia's gendering. This is the goal of this paper. We are particularly influenced by the STS literature that emphasizes 'the ways an infrastructure can structurally exclude some people (e.g. deaf, blind, or wheelchair-bound individuals) from purportedly "universal" services' (Plantin et al., 2016: 4). Taking a lead from Star (1999), we investigate the 'fundamentally relational' aspects of an information system that are often invisible and embedded within an 'installed base'. The concept of the 'installed base' refers to how infrastructures are always developed on top of other systems. In Wikipedia's case, these systems are predominantly developed from the styles and forms (and their accompanying epistemologies and logics) of two key projects: the Western encyclopedia and the free software movement. Recognizing the characteristics of Wikipedia's infrastructuring enables us to understand how Wikipedia not only extends the power relations that were prevalent in its installed base, but the ways in which Wikipedia's particular assemblage of human and non-human agents emphasize particular power relations while lessening others.

Instead of dealing with human and non-human elements separately, we connect different aspects of Wikipedia's infrastructure to the power relations, particularly in the realm of authority, that it enables. Wikipedia's infrastructure consists of the policies that guide contributions, and the norms and logics that determine what knowledge is and who the ideal knowers are. It also includes the material architecture of the system in the form of software code that governs the design of the system. In sum, these elements of Wikipedia's infrastructure consist of several interrelated dimensions. For the sake of clarity, we group them under the headings of architecture (wiki software and independent tools, as well as deeper layers, including search engines and web browsers); policies and laws (such as copyright and privacy); and norms and logics (epistemologies and norms for citation and verification, as well as norms of expertise).

Infrastructures produce power relations. In the case of Wikipedia, we are particularly interested in what it means to be an empowered and authoritative member of the Wikipedia community. We find that this requires particular types of skills and expertise, particular types of knowing that tend to be gendered. In the following sections we investigate the hidden logics embedded in Wikipedia's infrastructure that result from its installed base, before showing how those logics produce particular inequalities of expertise and authority at the level of code and policies.

The logics embedded in Wikipedia's installed base

What are the underlying logics about what knowledge is and who is best able to represent it on Wikipedia? Wikipedia epistemology is a foundational aspect of its infrastructure and is materialized in the policies and principles that guide work on the encyclopedia. Such logics determine what Wikipedia accepts and what it rejects. Logics about how knowledge is defined, who are the appropriate authors and experts, and which subjects are suitable for inclusion. This directly

influences how knowledge is produced. In order to understand how Wikipedia's epistemological position influences gender relations, we first look to the logics embedded in its installed base and move on to the particular policies that guide work on the encyclopedia, followed by the ways in which those policies are interpreted.

The encyclopedia is a particular genre of knowledge with historical roots in colonialism, primary education and technoscientific endeavors, that necessarily sets up limits about what can be added (and by whom it can be added) according to pre-defined norms and logics (Burke, 2012). Indeed, the envisioning of Wikipedia as the sum of all human knowledge is an historical continuation of earlier encyclopedias that set up a particular relationship between the expert and the layperson, the non-expert, or the learner, as the receivers of knowledge. The very definition of the encyclopedia is that it is comprehensive – that it holds a summary of *all* information from either all branches of knowledge or a particular branch. That encyclopedias describe themselves as representing all that is known reflects a particular understanding of knowledge that implicitly rejects the idea that some knowledge cannot (and perhaps should not) be known to all.

Encyclopedias have purported to represent *all* knowledge, but they have never sought to represent *everyone's* knowledge. Wikipedia, on the other hand, promises both in its invitation that 'anyone can edit' and to be working on representing the 'sum of all human knowledge'. Wikipedia's policies, however, reveal a significant conflict between these goals.

Wikipedia's content policies are centered around three core principles: neutral point of view (NPOV), no original research (NOR), and verifiability. NPOV demands that articles should be written without bias, by fairly and proportionately representing all significant views. The 'no original research' policy requires that Wikipedia editors do not publish original thought, and the verifiability policy determines that all material challenged or likely to be challenged must be attributed to a reliable source (Wikipedia, 2016).

Although NPOV is characteristic of previous encyclopedic epistemology, Wikipedia is unique in its profuse (and growing) use of in-text citations and sources for further reading, that accord with the verifiability logic. This verifiability policy is, in effect, a merging of the traditions and epistemologies of both scientific fields and the free and open source software movements. From science, verifiability draws from citation practices; from open source, it draws from the principle that the source should be transparent and that knowledge can be modularized.

Verifiability is not only a rule that applies to the way that statements are structured on Wikipedia; it is an ideological system that is constituted by a series of values, principles, policies and norms. Verifiability effectively defines the communicative acts that are possible within Wikipedia's sociotechnical system. To illustrate this, we might imagine an alternative perspective regarding verifiability that is concerned with equipping readers with the tools necessary to

verify whether the claim made in the text is, in fact, true according to their own observations. But Wikipedia policy stresses that content cannot be added unless it can point to a 'reliable source'.

The verifiability principle also sets up particular roles for editors, sources, readers and subjects on Wikipedia. Editors are banned from doing what is called 'original research' and must cite every claim made in an article that has been or could be challenged with a relevant source. The implication of this policy is that Wikipedia editors are positioned as merely passive aggregators of information already published by external sources. The authors of sources and citations, on the other hand, must be located outside the sphere of the encyclopedia; otherwise these authors will be regarded as having a conflict of interest. The source's independence from the phenomenon and from Wikipedia is an important criterion for determining their reliability and thus their acceptability for Wikipedia.

This policy focus, of adding only information that can be verified by a source deemed reliable by Wikipedia editors, has led to a series of unintended consequences. Primarily, it means that, in situations where knowledge is not verifiable or at least easily verifiable (within Wikipedia's definition of verifiability), it remains outside the corpus. In one example, a group of Wikipedians attempted to write about subjects from India, Malaysia and South Africa that were weakly represented in online sources by interviewing local experts in the 'oral citations project'. The project was, however, vehemently opposed on the English Wikipedia, where a majority of editors claimed that the subjects could not be independently verified and should therefore be removed from the encyclopedia (Gallert and Van der Velden, 2014).

When writing about women on Wikipedia, it is often the case that there are fewer published sources in the canon about women than men. As we noted above, for example, there is a dearth of content about women scientists on the encyclopedia and the ratio of male to female biographies is more pronounced on Wikipedia than Encyclopedia Britannica. The lack of citations or sources on a subject is used as a key reason for exclusion of new articles.[4] This is because such subjects are seen as not notable enough for inclusion if few sources considered reliable have written about them. In one example, a female Wikipedian, Zara Rahman, recounted how she had tried to add details about a long-neglected inventor, Hedy Lamarr, who was also a 1940s Hollywood star. When Rahman attempted to edit the article to reflect the significance of Lamarr's invention of early wireless technologies, she had her edits reverted on the grounds that Lamarr's acting career was noted as more significant by historical sources (Kleeman, 2015).

Wikipedia's infrastructure, in sum, extends and reinforces the biases of its installed base in the logics and principles that characterized the technoscientific project. We see this in the experiential and traditional (predominantly oral) knowledges that are rejected as a result, but we also see it in the emphasis on skills required to be a successful Wikipedian. Infrastructures are not neutral. Not only does Wikipedia's infrastructure reinforce existing power relations by

reaffirming the already legitimated power of scientists and academic professionals on Wikipedia, but it also introduces new, unequal power relations among editors who do not share the same levels of expertise and access to what are considered credible knowledge sources.

In this way, Wikipedia is extending the epistemologies of previous male-dominated technoscientific projects. As Haraway (1988: 581-83) writes, scientific objectivity is a 'truth regime'. Objectivity as it is practiced in male-dominated science is an illusion as there can be no 'infinite vision': it is a 'god trick'. From this perspective, all knowledges are local and situated; there can be no universal knowledge. Thus, the logics embedded in Wikipedia's installed base perpetuate science's provenance as a male domain as well as reflecting the legacies of colonialism and imperialism in the systematic exclusion of knowledge traditions from the Global South (see Connell, 2007; Harding, 2011; and footnote 1 here).

Wikipedia's software infrastructure

Wikipedia is maintained by a distributed and highly complex arrangement of non-human agents in the form of bots, software tools and bespoke code that automate many aspects of the encyclopedia's managerial structure. This is a significant departure from encyclopedias prior to Wikipedia and influences the power relations that result from this particularly coded material infrastructure.

Wikipedia's architecture is powered by MediaWiki, a free and open source software. Developed by individuals and organizations around the world, it is used to run sites and projects including, but not limited to, Wikipedia. The work is largely done by paid engineers at the Wikimedia Foundation headquarters in San Francisco. They invest a significant proportion of their donated funds into developing and maintaining the MediaWiki system.

MediaWiki employs an open API (application programming interface) that enables editors to plug in their own software applications when editing Wikipedia. Editors may also configure the MediaWiki software on their own terminals to employ a variety of 'skins' that change the appearance of the editing interface, or to enable tools such as an automated category suggestion add-on called 'HotCat'. Some editors have developed automated agents or bots that, as long as the Bot Approvals Group approves them, may perform automated editing tasks such as changing spellings of commonly misspelled words, adding templates or consolidating work. In addition, a host of non-wiki software is used by editors to coordinate activities including mailing lists, blogs, and internet relay chat channels, as well as tasks like server administration, vulnerability testing and donation processing. According to Geiger (2014), this 'bespoke code' is significant because many of the features and functionalities we take for granted in Wikipedia (such as the 'citation needed' tag) cannot be run without it.

The software that drives the site is widely distributed across the Wikipedia network, with the default software and content served by the Wikimedia Foundation's central servers, but with a wide range of software also deployed by

editors at the ends of the network. The manipulation of software and code by editors becomes critical to the literacy required to be an effective contributor. As a result, those whose editing is enhanced by the use of automated tools are much more effective contributors than those who use only the default tools (Ford and Geiger, 2012). The developers of software tools and developers and users of bespoke code are even more likely to be male than are editors. The proportion of female employees in the product and engineering team at the Wikimedia Foundation is about 25%. Some work has been done to address gender disparities in editing, such as a preference that editors can set so that the software uses gender neutral words whenever possible when mentioning them. The majority of female editors, however, do not advance to the stage where they can configure the system in this way.

Wikipedia's sociotechnical environment, its material infrastructure, is heavily mediated by code, so much so that the editing process is said to look more like a computer program than a draft of an encyclopedia entry. On the editing page, '[r]eams of code cascade down the page: curved, square and curly brackets, chevrons and underscores' (Kleeman, 2015). The ability to manipulate these coded objects is thus central to the discursive process by which knowledge claims are constructed and debated on Wikipedia. The travel of facts within this environment is influenced by the materiality of the code that produces the objects, tools, processes and conventions that drive Wikipedia.

The ideals that software reflects in its categorization of the world become a reality when such categorization becomes the common-sense way of framing particular social phenomena. When an infobox representing a biography on Wikipedia enables editors to choose only female or male categories, for example, this reinforces the gender binary. The ideal world as framed by the software reinforces a world categorized in this way. One of the effects of the abstraction by software algorithms and data models in rendering aspects of the world is that 'the world starts to structure itself in the image of the capta[5] and code – a self-fulfilling, recursive relationship develops' (Kitchin and Dodge, 2011: 41).

Furthermore, the software that has been used to power the editing process on Wikipedia has been criticized as too complex and technical for non-coders to use (Gardner, 2011; Halfaker et al., 2013). Although the Wikimedia Foundation invests significant resources into improving (and simplifying) the user experience, they have faced significant pushback from the older community of (mostly male) users, who are unhappy with the influx of those they regard as inexperienced editors entering the system. Indeed, the conservatism of this early Wikipedia community has been identified as the primary cause of Wikipedia's decline in editor numbers.

In the context of Wikipedia's adversarial and highly coded culture, authoritative editors are those who are able to either stabilize or destabilize an article by deploying a number of coded objects. In order to destabilize an article, editors might deploy objects such as warning tags or 'citation needed' tags, delete content added by others or nominate an article for deletion. In order to stabilize an article, editors must continue to add citations and sources, remove warning

tags, and continue to defend the article where it is being challenged. Additionally, editors may attempt to build alliances in order to provide the appearance of widespread support. What is achieved out of stabilization has been called 'consensus' in Wikipedia policy documents, but stabilization is the result achieved by the cessation of debate, not necessarily agreement by all stakeholders.

Those editors who are able to best perform the ideal identity of editors, passively collating and representing what already exists in independent sources, are those who prevail. Their power is a product of their ability to make a complex system work for their ideological goals. In this way, operating within the Wikipedia community resonates strongly with the broader masculine culture of computing and software work.[6]

Wikipedia's policy infrastructure

Not only is Wikipedia work highly technical; it is also highly legalistic. The term 'wiki lawyering' was used early on in the site's development to describe the kind of rhetoric employed by Wikipedians as they argue for the inclusion of particular content and sources by interpreting Wikipedia's core policies. The Wikipedian who is able to operate within the highly technical and legalistic framework of Wikipedia's infrastructure involves performing particular kinds of authority that, in turn, involves exercising particular kinds of power. Once more, this is consequential for the possibility of being a female Wikipedian.

Authority in Wikipedia is performed predominantly through the ability to engage in digital speech acts (Isin and Ruppert, 2015) by deploying both objects of policy and code. Policies are an important element of social interaction on Wikipedia (Bryant et al., 2005; Pentzold and Seidenglanz, 2006). Analysts have found that policies are used to appeal to authority in order to justify a contributor's changes to an article, and that policies provide a common resource for new users to learn about editing and behavioral conventions. There has been a significant rise in the number of policy and other administration pages on Wikipedia: between mid-2003 and late 2005, the number of administrative pages grew at a rate of nearly eight times that of main article pages (Viegas et al., 2007).

Wikipedia policies are numerous and complex, and encompass so many levels of authority that a user's relatively greater understanding of policy enables them to more effectively participate in debates in order to influence representation (Ford and Geiger, 2012). Wikipedia policy is an important feature in defining the rules by which participants delimit what is sayable and their complexity and ambiguity leads to power play among contributors as they try to gain control over an article (Kriplean et al., 2007: 172). Such power play is a significant element of Wikipedia culture and is disabling for female contributors who would rather not participate if participation requires strategies for either bolstering or undermining the positions of contributors (Menking and Erickson, 2015).

Another example of the adversarial culture of Wikipedia is in the extensive deletion debates that characterize its everyday practice. An average of 500 articles are deleted every week on English Wikipedia, either according to a process of deliberation amongst editors or a process in which an administrator unilaterally decides to delete an article according to criteria for speedy deletion. A study of article deletions on the English Wikipedia indicated that the process is managed by a relatively small number of longstanding users and that the majority of deleted articles are deleted under the criteria of 'no indication of importance' rather than for spam or copyright violations (Geiger and Ford, 2011). Schneider et al. (2013) found that familiarity with Wikipedia's policies and norms correlates with newcomers' ability to craft persuasive arguments, and that acceptable arguments employ community-appropriate rhetoric that demonstrates knowledge of policies and values. While such practices discourage all newcomers from participating in Wikipedia, performing this kind of technical expertise is typically associated with masculinity.

Conclusion: Wikipedia and the reconfiguration of expertise

Wikipedia is held up as the collaborative utopia: a world model of free, decentralized participatory democracy. Yet underneath this idealized image, an obdurate gender divide remains: the overwhelming majority of contributors are male. Those studying the issue have made important contributions to understanding this problem by examining the reasons for this inequity, including women's lack of technical skills and confidence, and female editors' experience of the adversarial culture of the Wikipedia community. Importantly, there have been several projects aimed at including more women and rectifying the male oriented content. Too often, however, the gender problem is still framed as a deficit in women.

The rich tradition of feminist STS has substantively foregrounded parallel arguments, yet, to date, it has been little drawn on in discussions of the gender divide in Wikipedia. An initial aim of this article, then, is to demonstrate how this earlier literature provides a broader conceptual grounding for contemporary analyses of Wikipedia as the sum of male knowledge.

Our central argument, however, it that the growing stream of infrastructure studies in STS provides the perfect tools for taking the gender analysis of Wikipedia to a further level. We have shown how being a Wikipedian – that is, being a member of the Wikipedia community – involves acquiring particular forms of sociotechnical expertise and authority that constitute the knowledge or epistemological infrastructure of Wikipedia. Beneath the rhetoric of the amateur, we found that a new form of expertise, and hence power, is being constituted, but that it is once again gender-coded as male.

Wikipedia is a knowledge institution governed by issues around power. Although presented as objective knowledge, the encyclopedia's frame necessarily limits what (and consequently who) can be included. Here we have shown that Wikipedia's material architecture, its platform, also presupposes specialist skills that advantage those already operating successfully within the

community. Those who are able to master Wikipedia's technocratic system of representation, with an emphasis on facts and other modular pieces of verifiable information, emerge as power brokers within this environment. Those who fail to master this system, either because their knowledge of the world does not fit with what Wikipedia recognizes as knowledge or because negative social interactions on the platform have led to their leaving it, will remain on Wikipedia's edges, unable to contribute and have their knowledge represented.

The infrastructure lens we have adopted here shines a light on the less visible sources of gender inequality. It reveals how the very identity of a Wikipedian, its habitus, still reflects the history of technoscience as an almost exclusively male province. Unless Wikipedia radically changes its own culture of knowledge production, women will remain on the edges; our knowledge will once again be marginalized.

Acknowledgments
We would like to thank the Oxford Internet Institute for providing both the intellectual context and the hospitality that enabled this collaboration.

Notes

[1] While our focus is on gender, others have provided evidence of content bias in Wikipedia at the level of knowledge area, where some subjects (particularly entertainment and popular culture) are covered in greater depth and detail than others (Salah et al., 2010) and knowledge about large parts of the developing world is unavailable (Gallert and van der Velden, 2015; Graham et al., 2014; Luyt, 2011). Unfortunately, although we do not mean to imply that women (or men) are a homogenous group, we could not pursue intersectionality here for lack of sufficient demographic data.
[2] A portmanteau of the terms 'bro' and 'programmer', 'brogrammer' has emerged in recent years to describe the rise of the testosterone-fuelled, power-wielding, white male programmer (see Macmillan, 2012).
[3] See Karasti et al. (2016) and Plantin et al. (2016).
[4] See Wyatt et al. (2016) on how Wikipedia's articles on schizophrenia are shaped by the ways in which the technical and policy architecture is employed and made to work for its editors.
[5] Kitchin and Dodge (2011: 5) define 'capta' as 'units that have been selected and harvested from the sum of all potential data… (for example), with respect to a person, data is everything that it is possible to know about that person, capta is what is selectively captured through measurement'
[6] See, for example, Margolis and Fisher (2002) for an analysis of masculine cultures of computing and Nafus (2011), who demonstrates that open source software projects reflect a similar phenomenon.

References

Balsamo A (1995) *Technologies of the Gendered Body: Reading Cyborg Women.* Durham: Duke University Press.

BBC News (2014) Wikipedia 'completely failed' to fix gender imbalance. BBC News [Video file], 8 August. Available at http://www.bbc.co.uk/news/business-28701772.

Berg AJ and Lie M (1995) Feminism and constructivism: Do artifacts have gender? *Science, Technology & Human Values* 20(3): 332-351.

Bryant SL, Forte A and Bruckman A (2005) Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In: *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, 1-10.

Burke P (2012) *A Social History of Knowledge, Volume II: From the* Encyclopédie *to* Wikipedia. Cambridge: Polity Press.

Cassell J (2011) A culture of editing wars. *The New York Times*, 2 February.

Cockburn C (1985) *Machinery of Dominance: Women, Men and Technical Know-How*. London: Pluto Press.

Connell R (2007) *Southern Theory: The Global Dynamics of Knowledge in Social Science*. Cambridge: Polity Press.

Collier B and Bear J (2012) Conflict, criticism, or confidence: An empirical examination of the gender gap in Wikipedia contributions. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 383-392.

Cowan RS (1979) From Virginia Dare to Virginia Slims: Women and technology in American life. *Technology and Culture* 20(1): 51-63.

Faulkner W (2001) The technology question in feminism: A view from feminist technology studies. *Women's Studies International Forum* 24(1): 79-95.

Ford H and Geiger R S (2012) Writing up rather than writing down: Becoming Wikipedia literate. In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (p. 16). ACM.

Ford H (2015) Fact Factories: Wikipedia and the power to represent (Doctoral dissertation, University of Oxford).

Gallert P and Van der Velden M (2015) Reliable sources for indigenous knowledge: Dissecting Wikipedia's Catch-22. In: Bidwell N and Winschiers-Theophilus H (eds) *At the Intersection of Indigenous and Traditional Knowledge and Technology Design*. Santa Rosa: Informing Science, 117-132.

Gardner S (2011) Nine reasons why women don't edit Wikipedia in their own words. In: *Sue Gardner's Blog*. Available at: https://suegardner.org/2011/02/19/nine-reasons-why-women-dont-edit-wikipedia-in-their-own-words/ (accessed 10 June 2016).

Geiger RS (2014) Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society* 17(3): 342-356.

Geiger RS and Ford H (2011) Participation in Wikipedia's article deletion processes. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration (pp. 201-202). ACM.

Glott R, Ghosh R, Schmidt P (2010) Wikipedia survey. Technical report, UNUMERIT, Maastricht, Netherlands. Available: https://web.archive.org/web/20131208221410/http://wikipediasurvey.org/docs/Wikipedia_Overview_15March2010-FINAL.pdf.

Graham M, Hogan B, Straumann RK, Medhat A (2014) Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers* 104(4): 746-64.

Halfaker A, Gieger RS, Morgan J, & Riedl J (2013) The rise and decline of an open collaboration system: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist* 57(5): 664-688.

Harding S (1986) *The Science Question in Feminism*. New York: Cornell University Press.

Harding S (ed) (2011) *The Postcolonial Science and Technology Studies Reader*. Durham, NC: Duke University Press.

Hargittai E and Shaw A (2015) Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society* 18(4): 424-442.

Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.

Haraway D (1997) *Modest_Witness@Second_Millennium.FemaleMan©_Meets_OncoMouse*®. New York: Routledge.

Henwood F (2000) From the woman question in technology to the technology question in feminism. Rethinking gender equality in IT education. *European Journal of Women's Studies* 7(2): 209-227.

Herring SC (2013) Discourse in Web 2.0: Familiar, reconfigured, and emergent. In Tannen D and Tester AM (eds) *Discourse 2.0: Language and New Media*. Washington, DC: Georgetown University Press, 1-25.

Hill B M and Shaw A (2013) The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS One* 8(6): e65782.

Hochschild A (1983) *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.

Iosub D, Laniado D, Castillo C et al. (2014) Emotions under discussion: Gender, status and communication in online collaboration. *PLoS One* 9(8): e104880

Isin E and Ruppert E (2015) *Being Digital Citizens*. London: Rowman & Littlefield International.

Karasti H, Millerand F, Hine C and Bowker GC (eds) (2016) *Knowledge Infrastructures*, Parts I-IV, Special issues of *Science and Technology Studies* 29.

Keller EF (1983) *A Feeling for the Organism: The Life and Work of Barbara McClintock.* San Francisco: Freeman.

Kitchin R and Dodge M (2011) *Code/Space: Software and Everyday Life*. Cambridge, MA: MIT Press.

Kleeman J (2015) The Wikipedia wars: Does it matter if our biggest source of knowledge is written by men? *New Statesman,* 22-28 May.

Koetsier J (2013) Free knowledge, freely delivered: Wikipedia Zero expands to 212M mobile users in developing nations*. Venture Beat,* 25 January.

Kriplean T, Beschastnikh I, McDonald DW and Golder SA (2007) Community, consensus, coercion, control: cs* w or how policy mediates mass participation. In: *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 167-176.

Landstrom C (2007) Queering feminist technology studies. *Feminist Theory* 8(1): 7-26.

Long JS and Frank Fox M (1995) Scientific careers: Universalism and particularism. *Annual Review of Sociology* 21: 45-71.

Luyt B (2011) The nature of historical representation on Wikipedia: Dominant or alterative historiography? *Journal of the American Society for Information Science and Technology* 62(6): 1058-1065.

Macmillan D (2012) The rise of the brogrammer. *Bloomberg*, March 2.

Margolis J and Fisher A (2002) *Unlocking the Clubhouse: Women in Computing*. Cambridge, MA: MIT Press.

McNeil M (ed) (1987) *Gender and Expertise*. London: Free Association Books.

Menking A and Erickson I (2015) The heart work of Wikipedia: Gendered, emotional labor in the world's largest online encyclopedia. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 207-210.

Nafus D (2011) 'Patches don't have gender': What is not open in open source software. *New Media & Society* 14(4): 669-683.

Niederer S and van Dijck J (2010) Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society* 12(8): 1368–1387.

Pentzold C and Seidenglanz S (2006) Foucault@ Wiki: first steps towards a conceptual framework for the analysis of Wiki discourses. In: *Proceedings of the 2006 International Symposium on Wikis*, 59-68.

Plantin J-C, Lagoze C, Edwards PN and Sandvig C (2016) Infrastructure studies meet platform studies in the age of Google and Facebook, *New Media & Society*, published before print, doi:10.1177/1461444816661553.

Reagle J and Rhue L (2011) Gender bias in Wikipedia and Britannica. *International Journal Of Communication* 5: 1138–1158.

Reagle J (2013) Free as in sexist? Free culture and the gender gap. *First Monday* 18(1): doi:10.5210/fm.v18i1.4291

Rose H (1983) Hand, brain, and heart: A feminist epistemology for the natural sciences. *Signs* 9(1): 73-90.

Rossiter M (1982) *Women Scientists in America, Struggles and Strategies to 1940*. Baltimore: John Hopkins University Press.

Rothschild J (ed) (1983) *Machina Ex Dea: Feminist Perspectives on Technology.* New York: Pergamon Press.

Salah AA, Gao C, Suchecki K and Scharnhorst A (2012) Need to categorize: A comparative look at the categories of universal decimal classification system and Wikipedia. *Leonardo* 45(1), 84-85.

Sayre A (1975) *Rosalind Franklin and DNA: A Vivid View of What It Is Like to Be a Gifted Woman in an Especially Male Profession*. New York: W.W. Norton & Co.

Schneider J, Samp K, Passant A and Decker S (2013) Arguments about deletion: How experience improves the acceptability of arguments in ad-hoc online task groups. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative work*, 1069-1080.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist* 43(3): 377-391.

Thomas G (1992) *A Position to Command Respect: Women and the Eleventh Britannica*. Metuchen, NJ: The Scarecrow Press.

Tkacz N (2014) *Wikipedia and the Politics of Openness*. Chicago: University of Chicago Press.

Turkle S (1984) *The Second Self: Computers and the Human Spirit.* London: Granada.

Viegas FB, Wattenberg M, Kriss J and Van Ham F (2007) Talk before you type: Coordination in Wikipedia. In: *HICSS 2007: 40th Annual Hawaii International Conference on System Sciences*, 78-78.

Wajcman J (1991) *Feminism Confronts Technology*. Cambridge: Polity Press.

Wajcman J (2004) *TechnoFeminism*, Cambridge: Polity Press.

Weinberger D (2011) *Too Big to Know*. New York: Basic Books.

Wikimedia Foundation (2011) Wikipedia editors survey. Available at: https://commons.wikimedia.org/wiki/File:Editor_Survey_Report_-_April_2011.pdf

Wikipedia (2004) WikiProject countering systematic bias. Available at https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Countering_systemic_bias

Wikipedia (2016) Core content policies. Available at: https://en.wikipedia.org/w/index.php?title=Wikipedia:Core_content_policies&oldid=714188784

Wyatt S, Harris A and Kelly S (2016) Controversy goes online: Schizophrenia genetics on Wikipedia. *Science & Technology Studies* 29(1): 13-29.

Author biographies

Heather Ford is University Academic Fellow at the University of Leeds School of Media and Communication. With a background in Internet rights activism, working with several organisations, including Creative Commons, Privacy International, the Association for Progressive Communications and Ushahidi, her research interests include issues around the governance of digital platforms, media power in networked information environments, and the design and politics of software platforms.

Judy Wajcman is the Anthony Giddens professor of sociology at the London School of Economics, and a visiting professor at the Oxford Internet Institute. She has published widely on the gender relations of technology. Her recent books include *Pressed for Time: The Acceleration of Life in Digital Capitalism* (2015) and

The Sociology of Speed: Digital, Organization, and Social Temporalities (2017, edited with N. Dodd).