

Production Misalignment: a Threat to Public Knowledge

Kaylea Champion

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Arts

University of Washington

2019

Reading Committee:

Benjamin Mako Hill, Chair

Adrienne Russell

Program Authorized to Offer Degree:  
Communication

©Copyright 2019  
Kaylea Champion

University of Washington

**Abstract**

Production Misalignment: a Threat to Public Knowledge

Kaylea Champion

Chair of the Supervisory Committee:

Benjamin Mako Hill

Department of Communication

On Wikipedia, when a high-interest topic is poorly-covered—either it is incomplete or contains inaccurate information—public knowledge is threatened. Contributors on Wikipedia are volunteers: they’re not assigned to track consumer demand, and they choose their own tasks. When contributor interest doesn’t align with consumer interest, the result is termed “underproduction”—some widely consumed materials are low quality. Past research has found competing explanations for what motivates volunteers to work on particular articles, including attempts to solve their own problems and supporting project goals. I theorize that social rewards explain task selection for moderate to high levels of experience, although this trend attenuates at the highest level of experience. Using a detailed longitudinal dataset, I find support for this theory in three ways. First, that although they are a minority of contributors, persistent contributors drive what gets produced. Second, as contributors persist, they are less likely to contribute to underproduced materials, but this trend flattens over time as predicted. Third, this pattern is weaker among contributors who do not create accounts.

## 1. INTRODUCTION

Wikipedia is the fifth-most popular site in the world, and often the first (and last) source of information people reach for. If you ask a digital assistant from Google, Amazon, or Apple about a topic, it will likely read the corresponding Wikipedia page to you. Wikipedia projects received 21 billion page views in October 2019 and existed in 292 different language editions. The largest edition, English, has nearly 6 million articles. When a subject is poorly covered in Wikipedia, the impact can be broadly felt—particularly when the topic is in high demand.

Wikipedia receives over 1.8 contributions per second from self-directed volunteers. This remarkable effort is an example of commons-based peer production: a public good emerging through a complex production process aided by technology tools, in which volunteers identify and complete tasks and then combine their work with the work of others. How do their choices of what to make correspond to what consumers want to use? In a market-driven system, producer supply and consumer demand are aligned via price—but in commons-based peer production, no such signal is available, and the alignment of supply and demand is not well-understood. Past investigations have pointed to social mechanisms as shaping what is produced, how, and by whom.

This project seeks to understand the emergence of a misalignment between what producers want to make and what consumers want to use. In §2, I examine what is known about how production decisions are made and how decision-making may vary among and within individuals depending on their level of experience and propose a hypothesis. I describe the setting, data, and analytical plan in §3 and share the results of a series of empirical tests in §4, with limitations as elaborated in §5. I discuss the implications of these findings in §6 and conclude in §7.

## 2. BACKGROUND

### 2.1. *Commons-based Peer Production*

Commons-based peer production is a term coined by Yochai Benkler to describe an emerging form of cooperative organization and work made possible by pervasive networks and digital technology. Peer-produced goods are non-rivalrous, developed through a collaborative volunteer process, and are an example of a collective public good (Fulk, Flanagin, Kalman, Monge, & Ryan,

1996; Ostrom, 1990). Wikipedia belongs to a class of collective public goods offering both connective and communal benefits (Fulk et al., 1996). As Benkler (2006) argued, networked modes of production create the potential for diminished separation between information producer and information consumer. However, the potential of these public goods is limited by the breadth, depth, and relevance of the topics which find a voice within them. One way to assess the success of a peer production project is to examine how well it meets the needs of those who use it. However, Warncke-Wang, Ranjan, Teevee, and Hecht (2015) found that over 40% of views to English Wikipedia were to relatively low-quality articles. In addition, consider Warncke-Wang et al.'s (2015) list of the topics disproportionately affected by this misalignment: countries, religion, LGBT topics, psychology, pop and rock music, internet and technology, comedy, and science fiction. I suggest that these neglected topics have a clear relevance to public affairs and/or represent interest in modern cultural materials. This misalignment of effort to demand suggests that significant information needs are going unmet.

Previous work has found that in general, popularity is a significant predictor of quality in Wikipedia, across languages as well as across multiple measures of quality (Lewoniewski, Węcel, & Abramowicz, 2017; Warncke-Wang et al., 2015). The Perfect Alignment Hypothesis (PAH) developed by Warncke-Wang et al. (2015) states that the goods in the highest demand would be the highest quality, while those goods in the lowest demand would be of the lowest quality. When quality is low relative to demand, these goods are characterized as *underproduced*. When quality is high relative to demand, then these goods are characterized as *overproduced*. Overproduction in the context of Wikipedia is not particularly harmful, but underproduction can have substantial negative consequences for information-seekers (Gorbatâi, 2011; Gorbatâi & Piskorski, 2012; Warncke-Wang et al., 2015).

## 2.2. *Experience and Task Selection*

In order to understand the sources for underproduction, it's important to consider why and how people participate in Wikipedia in the first place. The open nature of the public good allows free-riding by non-contributors, and the large scale of the project leads to difficulty in assessing whether one's own contribution makes a difference (Olson, 1965). Building a critical mass of contributors is difficult, given differential resources and interest level (Marwell & Oliver, 1993).

Information public goods projects are characterized by volunteer work and self-selection of tasks. Previous empirical work has found that peer production contributors report a range of goals: to create a free public good (Budhathoki & Haythornthwaite, 2013), to help other members of the production community (Wu, Gerlach, & Young, 2007), to enhance their own use of the public good (Krishnamurthy, Ou, & Tripathi, 2014; Meng & Wu, 2013), to enhance their reputation or improve their qualifications (Oreg & Nov, 2008), or to complete a class assignment (Coelho, Valente, Silva, & Hora, 2018).

The umbrella organizations for information public goods often operate as non-profits and articulate the goals of their project in terms consistent with the public interest. For example, the Wikimedia Foundation, which provides substantial support for the running of Wikipedia, presents itself with a home-page headlined: “Imagine a world in which every single human being can freely share in the sum of all knowledge.”<sup>1</sup>

How do volunteers select tasks, and how do their selections relate to the alignment of quality and demand? Past work on public goods contributors report a wide range of goals and interests, and also report variation in task selection with respect to experience. For example, in their examination of contributors to Open Street Map, Budhathoki and Haythornthwaite (2013) reported “serious” mappers (i.e. high volume contributors) more likely to be oriented to community recognition, while “casual” mappers reported wanting to contribute to a free and open project. By contrast, in Shah’s (2006) study, open source software contributors describe their first contributions as being directly related to personal skills, needs, and priorities, while longer-term participants reported conducting maintenance work for the general good of the project, as well as seeking out interesting challenges. Likewise, Bryant, Forte, and Bruckman (2005) found that Wikipedia editors initially reported participating first within their own areas of expertise, but that over time they sought to build the community and serve the public good.

However, Oreg and Nov (2008) found that contributor goals varied by the nature of the work produced. Software contributors (i.e. the group described in Shah (2006)) reported being driven to seek out tasks that would build their reputation, while content contributors reported being drawn to work that served the public good. Another site-related difference may be driven by group size and perception of one’s own relative capabilities and competencies: not only due to concerns about monitoring the contribution levels of others

---

<sup>1</sup><https://wikimediafoundation.org/>, viewed on 10/15/2018.

(Marwell & Oliver, 1993), but also because contributors may be driven by the norm of reciprocity (Connolly & Thorn, 1990; Fulk et al., 1996).

To develop a hypothesis about the overall trend for contributor task selection, this analysis examines evidence in three categories: initial contributors, experienced contributors, and the highest-volume contributors. Studies of new contributors have revealed that individuals initially select tasks which relate to their own immediate needs or areas of knowledge and which appear to be within their skill range (Bryant et al., 2005; Preece & Shneiderman, 2009). Some preliminary preferences in task selection may persist regardless of accumulated experience (e.g. the geographical biases in the case of Open Street Map may reflect the actual homes of contributors (Thebault-Spieker, Hecht, & Terveen, 2018)). Given that most articles are aligned, it seems most likely that initial contributions would be to aligned articles.

Contributions may at first go unnoticed, or receive some social rewards such as recognition from other contributors (Beschastnikh, McDonald, Zachry, Kriplean, & Borning, 2009; Oreg & Nov, 2008) and enjoyable experiences of group collaboration (Bryant et al., 2005; Budhathoki & Haythornthwaite, 2013; Meng & Wu, 2013). Tellingly, Piskorski and Gorbatâi (2017) found that in-group social rewards of the kind that contributors receive from producing information public goods are more likely to be found in areas of greater social density, in which like-minded editors develop familiarity and affinity with one another in working on articles within a fixed topical range. Piskorski and Gorbatâi (2017) delimit these topical ranges using the multi-dimensional category scheme of Wikipedia itself, and use network analysis of contributors who make revisions to the same article in temporal proximity to one another to assess associated social density. These dense networks of experienced contributors suggest higher quality, and fixed topical range suggests niche topics, which may have average lower demand: these are the two factors (high quality, low demand) which define overproduced goods.

However, prior work gives reason to believe that a trend driven by social reward factors is limited. Instead, the influence of social rewards may diminish as they accumulate (Bénabou & Tirole, 2006; Willer, 2009). Peer production organizations often describe their projects in terms oriented toward the public good, and as described above, some longer-term participants in empirical studies have reported a shift toward taking up efforts for the good of the project. To the degree that the most prolific participants may take on public interest tasks rather than their own topical preferences, one would expect

to find that the most experienced contributors are engaged in the improvement of underproduced goods, which reach relatively more people and are relatively worse quality.

This analysis considers popularity and quality at a given point in time, and examines individual contributors with respect to their own experience level at the moment of contribution. This is important because individuals make an initial decision to join and then make numerous decisions to persist in contributing, some of them thousands of times over the course of years. Having examined how one might expect contributors to select tasks as their experience accumulates, as well as trends that may drive the choices of very high-volume contributors, what remains is to connect these pieces to suggest an overall shape for how alignment changes with respect to accumulated experience.

Synthesizing from the previous work and analysis above, I predict contributors will initially contribute to aligned materials, followed by a movement toward overproduced materials as social rewards pattern their contribution, followed by some limitation or turning point driving toward underproduced materials in the tail. Visually, this movement in one direction and then another suggests a U shape, and if overproduction is represented as movement in the positive direction and underproduction in the negative direction, these trends together suggest *H1: a downward-facing, U-shaped relationship between participant experience and article alignment.*

### 2.3. Identifiability

Another factor may change the influence of experience: the establishment of a stable identifier. Receiving in-group social rewards for a contribution is dependent on some level of identifiability—collaborators need to know to whom they should direct their personal responses, and contributions need to be affiliated with some stable identifier to allow for attribution. This connection between social reward and identifiability is supported by past research suggesting that the decision to create an account may be driven by a desire to obtain greater feedback or recognition within the community (Forte, Andalibi, & Greenstadt, 2017).

When a contribution is made without an account, the contribution records for the article shows only the IP address of the individual. An IP address is associated with a geographical location at a given point in time, and can vary as an individual moves through their day. Hence this analysis of contributors

who participate without an account is limited, since it does not track individuals across multiple IP addresses or assess whether they also have an account. However, incorporating IP-based contributors into this analysis also to some extent controls for social interactions. Based on these arguments and past results, I hypothesize that in-group social rewards will be much less salient for contributors without accounts. As a result, I hypothesize that *(H2) experience level will have no significant effect on the article selection patterns of contributors without accounts.*

### 3. RESEARCH DESIGN

#### 3.1. Empirical Setting

This study is conducted in the context of Wikipedia, which defines itself as “a multilingual, web-based, free-content encyclopedia project supported by the Wikimedia Foundation and based on a model of openly editable content.”<sup>2</sup> The text and design of the encyclopedia are freely editable: with some few exceptions, if someone wants to contribute to Wikipedia, clicking on a tab labeled ‘Edit’ will bring up an interface for making a revision to the existing text. Saving this revised version will allow the change to be instantly visible, although this process varies somewhat by language edition: some languages apply filters (such as checking for profanity or tell-tale signs of spam), and some tend to scrutinize the contributions of new or unregistered editors more highly.

Although the content of its many articles are the primary way that the public interacts with Wikipedia, a complex community operates just behind the scenes, engaged in such tasks as deliberating article content and site features, coaching and encouraging one another, organizing and analyzing the work to be done, developing rules and enforcing them, and patrolling through changes to reverse vandalism or inadvertent damage to pages. The Wikimedia Foundation makes a wide range of rich data available, including the content of every revision to every article going back to the origin of Wikipedia in 2001, as well as the number of views to each article going back to December 2007, and the dataset analyzed in this project is drawn from this public data.

---

<sup>2</sup><https://en.wikipedia.org/wiki/Wikipedia:About>

### 3.2. *Sample*

The dataset consists of all revisions made to Wikipedia since its inception, up until March 2018 (when this work began). To develop my sample, I first considered the wide range of activities which occur on Wikipedia. Given the objective to understand trends affecting human article editing over the course of accumulated experience, I discard all data related to accounts known to be bots, which I identify by scraping a Wikipedia webpage listing current registered bots, and from a dataset produced and released as part of Geiger and Halfaker (2017).

Editing activities follow a highly skewed, long-tailed distribution—a majority of contributors edit Wikipedia one time and never return, while a highly active editor may contribute hundreds or even thousands of revisions in a given month. In order to obtain data which incorporates the experiences of both high-volume and low-volume contributors, I drew this sample first by breaking editors into strata on the basis of their total accumulated number of revisions to articles, from the inception of Wikipedia to March 2018, sized by powers of two (i.e. the first strata contains all editors who made one article revision in the history of Wikipedia; the second strata contains all editors who made two article revisions, the third contains editors who made three or four revisions, and so on). I drew a random sample of 5% (five percent) of all editors within each experience strata and then extracted all revisions made by that editor.

I excluded any revisions made to the simple functional pages called ‘redirects’ (Hill & Shaw, 2014). Redirects are not articles per se, although they appear in logs as if they were articles—they are used to manage the fact that a single article may be known by multiple names or spellings. I also excluded any revision that was later deleted, and any for which the quality score could not be obtained. Because Wikipedia contains a wide range of content, including discussion pages, socializing, and personal pages, I also eliminated any revisions made to non-article pages, which reduced the sample to 9,671,242 revisions. I filtered disambiguation pages and “List of” pages: both are composed of lists of links but intentionally have very little content; this dropped the sample to 9,319,960 revisions. I removed any revisions made to pages which were moved after the edit of interest was made, because the moving procedure renames and collapses the history of the page in ways that obscure the quality and popularity of the page at the time it was edited (Hill & Shaw, 2015).

I also removed identity reverts: quality control actions which ‘undo’ some

previous revision or revisions and restore the article to a previous state without adding any new content. Reverts of this kind are often made through tools which look for patterns characteristic of vandalism and then route work to vandal-fighters. Identity reverts are detected by examining the cryptographic hash of a page after an edit action and then looking forward and backward 15 steps in the revision history. Removing reverts reduced the sample to 8,408,452 revisions. Finally, to avoid repeated measures issues I also removed any revisions where an editor made multiple revisions to the same article within the same month; this dropped the sample to a final total of 5,175,649 revisions.

### 3.3. Variables

I used the *revision* as the unit of analysis. I operationalize editor experience as *revision count*: the number of revisions the editor has made up to and including the current revision, and treat it as an independent variable. In order to consider the influence of receiving social rewards, I also introduce an independent factor variable for contribution without use of an account: *IP-based*. Those who create an account are termed *account-based*.

	Minimum	Maximum	Mean	Median	Std Dev
View Count	1	10800609	1614	1462	10
Revision Count	1	832281	47004	1115	119666

Table 1: Descriptive statistics of key measures in our sample.

The dependent variable is *alignment*, calculated as the residual of a “perfect” relationship between quality and popularity: that is, as defined in the Perfect Alignment Hypothesis (Warncke-Wang et al., 2015). I count views at the month level, and likewise calculate quality at the month level using the ORES quality measure, which focuses on objective features (such as the presence of references, section headings, and inter-wiki links) (Halfaker, Morgan, Sarabadani, & Wight, 2016).

ORES was trained by Wikipedians using quality classifications from 2015 and 2016, and uses the six quality levels defined in Wikipedia: Stub, Start, C, B, Good Article, and Featured Article. ORES produces a continuous prediction value from 0 to 1 for each quality grade, and the predictions for all six possibilities sum to one. In order to produce a single measure of quality, I apply a simple weighting corresponding to the ordering of these factors: the

lowest quality is weighted 1, and the highest quality 6, which results in a combined quality score with a theoretical range of 1 to 6; in my sample I observe a range of 1.04 to 5.89. As a robustness check on this alternate quality measure, I calculated the Pearson's correlation between the Wikipedian-assigned quality and the ORES-detected quality for the month of data analyzed by Warncke-Wang et al. (2015) to be .696.

I then used linear regression to assess the relationship between views and quality. Since overall usage levels of Wikipedia have varied over time, I also added a factor for the month and year as a control. I examined whether a polynomial model might improve model fit but found that the substantive result was unchanged:  $R^2$  of the linear model and the quadratic model are both 0.378.

The dependent variable, *alignment*, is the residual from the model predicting the quality of each article in each month, given its number of views. When the residual is near zero, the article has views and quality in approximate alignment. A negative residual indicates that the number of views predicted higher quality than what ORES observed. A positive residual indicates that the number of views predicted lower quality than what ORES observed. A difference of 1 indicates a difference of one quality class, and a range of -5 to 5 is possible.

I examined extreme values to validate this interpretation. For example, one of the highest negative residual values (i.e. most underproduced) articles that was edited in the random sample was to the article “Death of Osama bin Laden” in the month when he was killed. That month, the article received over 140,000 views, but was on average only Stub quality. Other extremely underproduced articles concerned people, media, and events during times when they might be most expected to be in the news: actors when a breakout role was aired, sports events while they were underway, and so on.

By contrast, overproduced topics in the sample reflect specialist interest or knowledge: one such overproduced topic was “History of FK Sarajevo”, that is, the Sarajevo football club, which received fewer than 100 views in April 2017 but was gauged on average to be Featured Article quality. Other extremely overproduced articles concerned such topics as small battles of broader wars, specific roads, small parks and islands, and specialized scientific topics. The final revision sample contains 5,175,649 revisions made by 1,116,389 uniquely identified editors—931,868 accounts, and 184,521 IP addresses.

### 3.4. Analytical Process

I employed two regression models, each providing some insight into both *H1* and *H2*. The first model uses a linear mixed-effects regression model that includes a random intercept term for users. Although this model attempts to correct for repeated measures of users, it might fail to capture the fact that experienced users might edit different articles than less experienced users in ways that simply reflect underlying differences in the types of users who go on to become more and less experienced (Panciera, Halfaker, & Terveen, 2009). To address this, I also fit a second model, a linear model with user-level fixed effects to capture the within-editor effect of change in experience. Because there is no within-editor variation in whether a given individual is identified by their account or by their IP address, I cannot include *IP-based* as a measure in the within-editor model; instead, I fit the model on the whole dataset and on the two subsets by identification type.

## 4. FINDINGS

I used a linear mixed-effects model to assess between-editor trends with current experience as the predictor, and allowed for a random effect for the final experience level of the contributor. A hierarchical approach is suitable for analyzing data which has some nested structure of repeated measures, as in the case where multiple revisions are made by a single individual.

The results from the mixed-effects model are shown in Table 2.<sup>3</sup> I find that the predictors for the linear effect of experience are statistically significant for both account-based and IP-based contributors. The coefficients are numerically small, and already on a log scale, such that the first-order coefficient for experience (0.0221) indicates that for a 1% increase in the number of revisions an individual has performed, the model predicts a shift of  $2.21 \times 10^{-4}$  toward overproduced articles. For editors contributing without an account, the second-order interaction coefficient is not statistically significant, although the linear effect is. Given a range of alignment from -5 (underproduction) to 5 (overproduction), these effects appear to be a small, as you see in the marginal effects plot in Figure 1. For example, a prototypical editor with 25 revisions is

---

<sup>3</sup>I performed an ANOVA to compare this model to a null model using only the mean and a random effect for each user in order to assess statistical significance, which yielded a highly significant  $p < .0001$  in a  $\chi^2$  test and reduced the BIC by 709, a substantial improvement in fit.

Between-Person Model:Alignment	
Intercept	−0.0566 [−0.0606; −0.0525]
Revision Count (ln)	0.0221 [0.0203; 0.0239]
Revision Count (ln <sup>2</sup> )	−0.0012 [−0.0013; −0.0011]
Editor was IP-based	0.0238 [0.0193; 0.0282]
Revision Count (ln) and IP-based	−0.0104 [−0.0132; −0.0076]
Revision Count (ln <sup>2</sup> ) and IP-based	0.0001 [−0.0003; 0.0006]
AIC	12360321.9868
BIC	12360429.6626
Log Likelihood	-6180152.9934
Num. obs.	5175649
Num. groups: editor.id	1116389
Var: editor.id (Intercept)	0.0822
Var: Residual	0.6099

Table 2: Results of a mixed-effects model to examine between-person change in the average alignment level of articles selected for editing. Bracketed values indicate a 95% confidence interval.

predicted as editing articles which score as 0.0022, while an editor with 1000 revisions is predicted as editing articles which score as 0.039.

That said, to assess the effect size of this finding, I consider that the large scale of Wikipedia can magnify even small changes. To understand the human impact of changes of this magnitude, I consider a hypothetical situation: what if everyone edited like the person with 25, versus what if everyone edited like the person with 1000? Given that the measure of interest, production alignment, is essentially utilitarian, I suggest interpreting effect using the notion of beneficiaries. If an editor improves an article with many readers, versus an article with only a few readers, there were more beneficiaries of the former effort than the latter. Recollecting that the alignment measure relates quality and views, for this hypothetical situation, then, I consider how choosing to edit articles which vary by relatively small amounts of alignment might be associated with different numbers of beneficiaries.

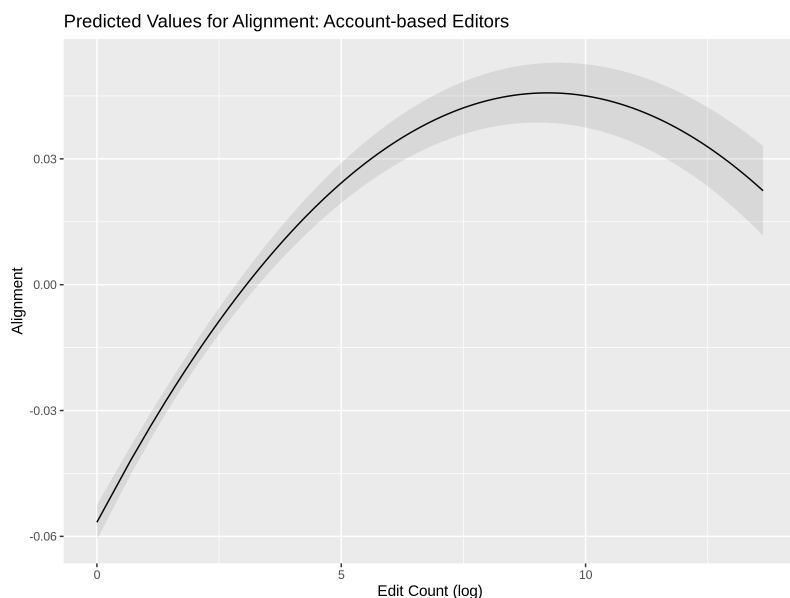


Figure 1: Between-person marginal effect of having higher experience on the average alignment of an article selected for editing. Increasing values indicate overproduction, i.e. high-quality but under-viewed topics.

A quality score of 2.5 is the approximate average of the articles edited in the sample, representing articles somewhere between “Start” class and “C” class quality. The Wikipedia quality guide says a Start-level article “Provides some meaningful content, but most readers will need more.” while a “C” class article is “Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study.”<sup>4</sup>

These models predict that the prototypical quality 2.5 article of the type that tends to be selected by an editor with 25 revisions may receive approximately 69 more views than the article chosen by the editor with 1000 revisions of accumulated experience. Although this may seem like a moderate impact, at the scale of Wikipedia, changes in average behavior are substantially magnified. In August 2019, English Wikipedia received about 4,500,000 human revisions to “content” areas. If each of these revisions reached 69 fewer individuals, these models predict that the collective effort of Wikipedians that month was directed into locations that were viewed about 310,500,000 fewer times. This number of views accounts for about 3.5% of the 8,960,173,622 views to Wikipedia content pages during this time.

<sup>4</sup>[https://en.wikipedia.org/wiki/Wikipedia:Content\\_assessment](https://en.wikipedia.org/wiki/Wikipedia:Content_assessment)

	Alignment: All	Alignment: Account	Alignment: IP
Revision Count (ln)	0.0104 [0.0082; 0.0125]	0.0234 [0.0202; 0.0266]	-0.0001 [-0.0038; 0.0035]
Revision Count (ln <sup>2</sup> )	-0.0006 [-0.0007; -0.0005]	-0.0013 [-0.0015; -0.0011]	-0.0000 [-0.0005; 0.0005]
Num. obs.	4697644	3339611	1358033
R <sup>2</sup> (full model)	0.2292	0.1405	0.4606
R <sup>2</sup> (proj model)	0.0000	0.0001	0.0000

Table 3: Results of cross-sectional panel modeling to examine within-person change in the alignment of articles selected. Bracketed values indicate a 95% confidence interval. Projected  $R^2$  indicates the very poor model fit that would result if I modeled this data with random effects for experience level but without fixed effects for the individual.

In addition to predicting impact on the number of people who benefit from volunteer efforts, the between-person model also offers insight into the value of IP-based contributors. We see negative coefficients in Table 2 for the interaction between revision count and IP-based contributors in the linear term, a statistically insignificant coefficient for the squared term, and a positive coefficient for the main effect of being IP-based. This result suggest that although the starting position of IP-based contributors is higher (i.e. given that the Intercept is slightly negative, this places IP-based contributors closer to the aligned zero point, and also initially closer to overproduced topics) than account-based contributors, the effect of increasing experience are attenuated among IP-based editors.

The mixed-effects between-person model described above has an inter-class correlation (ICC) of 0.119. That is, 11.9% of the variation in the sample is located within individuals. I therefore turn to the model of within-person differences, making use of cross-sectional panel modeling. Examining within-person change also allows detection of potential “survival bias”: high-volume contributors may be systematically different from casual contributors, so I test if these findings are consistent within individuals.

Because these models seek to estimate within-user change, this model is fit on the 4,697,644 revisions made by the 638,384 contributors who made more than one revision in the same stratified random sample used to fit the between-person model described previously. The results of the within-person model for alignment are reported in Table 3.

The within-person results are consistent with the between-person results,

although the relative thinness of the sample for high-volume contributors and the low ICC suggests interpreting these results with caution. The within person trend is positive in the linear term and negative in the squared term, suggesting a downward-sloping curve, similar to the between person trend; the trend for IP-based editors is not significant. This result suggests that IP-based editors do not generally change in their preferences as they accumulate experience.

This result is not surprising given the ICC finding placing the bulk of the variation as existing between, not within, people. This analysis suggests a phenomenon of survivorship bias with respect to alignment: while editors' preferences with respect to article alignment are relatively stable, at larger scale these contributions tend toward overproduced articles, but this effect is small compared to the differences between people.

## 5. LIMITATIONS

This study is limited in several important ways. First, these measures of quality and popularity are taken at the per-month level, but views and quality are at times more sporadic and taking a month view levels out this variation. These measures therefore represent the overall quality of an article within that month, and its overall viewership, but not its popularity or quality at the exact moment of editing.

Further, this analysis does not control for the fact that at some IP-based revisions are likely to be authored by experienced Wikipedians. This proportion may be significant, as suggested in Jackson, Crowston, and Østerlund's (2018) study of citizen science participants. Editors who typically work under their own name or who have built up social connections associated with a pseudonym may prefer to keep some dimensions of their identity, interests, and online production efforts private. Thus it may be that IP-based editor data is a conservative count of revisions. Wikipedia treats the IP address of registered contributors as private data and does not make it generally available to researchers. However, this makes these findings conservative: the distinct trend in the behavior of experienced users is limited to those who are in some sense enacting publicly their level of experience. In addition, although this analysis measures the revision count of IP addresses, this measure is likely to be noisy for multiple reasons: IP addresses assigned to a given home or business change may change periodically, and people may bounce among several addresses; hence, this analysis cannot consistently associate individuals with

IP addresses.

I also mitigate but do not fully address a risk described by the Wikipedia editing community when Warncke-Wang et al. (2015) was published. Community members stated that one explanation for underproduction is that some articles are more difficult to write than others, and in fact that difficulty may be systematically distributed in a way that coincides with the ways in which people search for and consume articles. For example, a general topic may be a more common search term than a niche topic. For producers, the effort involved to thoroughly develop a high-quality article about a general topic may also substantially larger than the effort involved to develop a high-quality article on a fairly niche topic, at least with respect to human judgment of quality.<sup>5</sup>

Thus it is possible that the same factor (conceptual generality) drives both high popularity and low quality. The use of ORES as a quality measure mitigates this risk in part because ORES measures quality based on structural features, independent of conceptual qualities such as completeness. However, this approach does not address the fact that, as a tertiary source, Wikipedia is limited by the availability of reliable published information. This means the difficulty of writing a high-quality article about a high-interest but under-reported topic may be more difficult than writing about widely-documented topics.

## 6. DISCUSSION

This paper offers insight into longitudinal changes in production behaviors and the impacts of these changes. These findings are consistent with *H1*, a movement from aligned, to over-produced, then ultimately to under-produced materials; a downward-facing U-shaped relationship between experience and contributions to underproduced goods. This trend resides both between and within individuals, but that the within-person trend is limited, which indicates a survival bias: those who edit overproduced articles are also those most likely to contribute moderate to high numbers of revisions, and those who contribute to underproduced materials are also those most likely to contribute very high numbers of revisions. These results also provide support for *H2*: the effect of experience will be attenuated in contributors without accounts. The between-person model predicts a diminished effect size for experience when

---

<sup>5</sup>[https://en.wikipedia.org/wiki/Wikipedia:Wikipedia\\_Signpost/2015-04-29/Recent\\_research](https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-04-29/Recent_research)

contributors are IP-based, and the within-person model shows no statistically significant change as experience increases.

Examining between person versus within person trends gives some insight into how underproduction emerges: not only do individuals turn away from improving underproduced articles over time, the people who persist in contributing and therefore contribute higher volumes of materials are those who do not tend to contribute to underproduced materials. These two trends taken together would tend to promote and sustain underproduction. In addition, this analysis suggests that IP-based individuals offer unique value as being willing to contribute to underproduced materials.

Although these results do not supply complete details of the mechanisms which may drive this trend, these findings are consistent with a feedback loop: given that social rewards are more available to those working in high-density topics and within a fixed topical range, these editors are those more likely to perceive themselves to be part of a group, to reward one another reciprocally, and persist. At very high volumes, the trend becomes less stable and although the data is thinner here, this evidence suggests that these maximal contributors are drawn toward underproduced materials.

In addition, given Warncke-Wang et al. (2015)'s previous results which suggest that underproduced articles are more likely to be related to current events, popular culture, or controversial topics, underproduced topics may be in some way less appealing to moderate to high volume contributors. Perhaps the topics are not associated with reciprocating social groups, are considered lower status, or are high-conflict in a way which is discouraging, or have some other trait such as cultural taboos that makes them less likely to encourage the development of social groups to encourage and socialize newcomers.

This analysis finds only limited evidence for reversals in underproduction trends over time, neither between nor within individuals. This is consistent with, and expands on the findings in Panciera et al. (2009) that the highest-volume contributors are "Born Not Made". Contributors tend to be relatively invariant in the alignment of the articles they work on, and that moderate to high-quantity contributors display distinct tendencies. Given the relatively low incidence rate of the highest-quantity contributors, these results suggest that communicative and communal public good production communities may need to take explicit steps to close production gaps. As Gorbatai (2014) found, signals of popular attention may draw production activity from experienced participants. Given the power of social forces to shape produc-

tion behaviors that this analysis suggests, it may be that accepted lines of action within the community and emic vocabulary can be invoked to address gaps as they emerge.

Alignment of goods is ultimately about their fitness for purpose: whether what's provided is a strong match for what's wanted. Information goods that do not match the interests or meet the needs of information seekers may ultimately benefit successively fewer people over time, no matter how large their store of unaccessed value. Hence findings of underproduction are concerning for the ongoing relevance and ultimately the sustainability of information public goods.

Past work has found that contribution levels to Wikipedia have declined, and that this change is attributed to increased rejection of newcomers and automated rejection responses to mistakes rather than human socializing responses (Halfaker, Geiger, Morgan, & Riedl, 2013). Given that newcomers are more likely to edit underproduced topics, this analysis gives reason to suspect that declines in newcomers would be associated with declines in alignment, and makes interventions which succeed in retaining diverse newcomers all the more important, e.g. the Wikipedia Teahouse (Morgan & Halfaker, 2018). However, given this result that low-quantity and non-accountholding contributors are more likely to contribute to aligned and underproduced topics, it may be that maintaining low barriers to entry for these very casual participants is also a useful strategy for countering underproduction. In future projects, I will be investigating how an underproduction measure may translate to diverse production contexts and how additional factors may predict or remediate production gaps.

## 7. CONCLUSION

This study joins an emerging body of work which examines the production implications of the social dynamics of a peer production platform. Although previous research suggests that the underproduction of peer produced information goods is widespread and important, this prior work has not attempted to explain its sources. Building on this research, I advance a theory that experience contributes to shifts in production behavior, with a mechanism of social and communicative feedback loops. I test this theory and find support: individuals tend to choose progressively less aligned articles as they accumulate experience, and more importantly, those people who tend to edit less aligned articles are also those who tend to contribute at high volumes. Contributing

from an IP is an important factor, mitigating against this trend of neglecting high-demand articles.

Work of prior investigators suggested that shifting motivation may reflect a source for behavioral change; however these shifts are reflected primarily in macro-level changes: most of the observed change resides at the between editor level; that is, a change driven by the task selection patterns of those who persist. Within-individual changes are modest. Over time, these findings predict that the ongoing efforts of the dedicated volunteers who produce Wikipedia will tend to reach fewer and fewer viewers. IP-based editors (whose within-person trend and between-person trend both suggest a tendency to edit underproduced articles) and newcomers (because their average alignment for early contributions is lower), provide an important counter to these trends. Communities which seek to produce goods which meet the demands of their consumers may need to find additional mechanisms to communicate production gaps and to directly encourage production of highly desired but low quality goods. However, given the relatively flat within-person trend and evidence for survivorship bias, progress may not lie in changing contributor behavior, but rather in supporting the recruitment, engagement, and retention of contributors with diverse preferences.

Online information-seeking includes diverse uses of Wikipedia, since not only individuals but also numerous platforms draw upon Wikipedia to support collective sensemaking. Wikipedia is a source for background knowledge on pressing public issues and a source of fact-checking for current events, recurrent myths, and misunderstandings. Gaps in Wikipedia are ultimately a threat to public knowledge. Understanding what factors may diminish the engagement of contributors willing to improve poor-quality articles on high-demand topics would protect the continuing value of this precious public good.

## REFERENCES

- Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), 1652–1678. doi:10.1257/aer.96.5.1652
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.
- Beschastnikh, I., McDonald, D. W., Zachry, M., Kriplean, T., & Borning, A. (2009). Promoting Quality in Wikipedia through Enculturation, "Supporting the amateur" Workshop: GROUP'09.

- Bryant, S. L., Forte, A., & 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* (pp. 1–10). GROUP '05. doi:[10.1145/1099203.1099205](https://doi.org/10.1145/1099203.1099205)
- Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist*, *57*(5), 548–575. doi:[10.1177/0002764212469364](https://doi.org/10.1177/0002764212469364)
- Coelho, J., Valente, M. T., Silva, L. L., & Hora, A. (2018). Why We Engage in FLOSS: Answers from Core Developers. In *Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering*. doi:[10.1145/3195836.3195848](https://doi.org/10.1145/3195836.3195848)
- Connolly, T., & Thorn, B. K. (1990). Discretionary databases: Theory, data, and implications. In J. Fulk & C. Steinfield (Eds.), *Organizations and communication technology* (pp. 219–233). Newbury Park, CA: Sage.
- Forte, A., Andalibi, N., & Greenstadt, R. (2017). Privacy, anonymity, and perceived risk in open collaboration: A study of Tor users and Wikipedians. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1800–1811). CSCW '17. doi:[10.1145/2998181.2998273](https://doi.org/10.1145/2998181.2998273)
- Fulk, J., Flanagan, A. J., Kalman, M. E., Monge, P. R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, *6*(1), 60–87. doi:[10.1111/j.1468-2885.1996.tb00120.x](https://doi.org/10.1111/j.1468-2885.1996.tb00120.x)
- Geiger, R. S., & Halfaker, A. (2017). Operationalizing Conflict and Cooperation between Automated Software Agents in Wikipedia: A Replication and Expansion of ‘Even Good Bots Fight’. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–33. doi:[10.1145/3134684](https://doi.org/10.1145/3134684)
- Gorbatâi, A. D. (2011). Exploring Underproduction in Wikipedia. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 205–206). WikiSym '11. doi:[10.1145/2038558.2038595](https://doi.org/10.1145/2038558.2038595)
- Gorbatâi, A. D. (2014). *The paradox of novice contributions to collective production: Evidence from Wikipedia* (SSRN Scholarly Paper No. ID 1949327). Social Science Research Network. Rochester, NY. Retrieved August 29, 2019, from <https://papers.ssrn.com/abstract=1949327>
- Gorbatâi, A. D., & Piskorski, M. J. (2012). Social structure of contributions to wikipedia. *Harvard Business School*. Retrieved September, 23, 2013. Retrieved from <http://www.academia.edu/download/3445150/AndreeaGorbattai.pdf>
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia’s reaction to

- popularity is causing its decline. *American Behavioral Scientist*, 57(5), 664–688. doi:10.1177/0002764212469365
- Halfaker, A., Morgan, J., Sarabadani, A., & Wight, A. (2016). *ORES: Facilitating re-mediation of Wikipedia's socio-technical problems*. Wikimedia Research. Retrieved from [https://meta.wikimedia.org/wiki/Research:ORES:\\_Facilitating\\_re-mediation\\_of\\_Wikipedia%27s\\_socio-technical\\_problems](https://meta.wikimedia.org/wiki/Research:ORES:_Facilitating_re-mediation_of_Wikipedia%27s_socio-technical_problems)
- Hill, B. M., & Shaw, A. (2014). Consider the redirect: A missing dimension of Wikipedia research. In *Proceedings of The International Symposium on Open Collaboration* (28:1–28:4). OpenSym '14. doi:10.1145/2641580.2641616
- Hill, B. M., & Shaw, A. (2015). Page protection: Another missing dimension of Wikipedia research. In *Proceedings of the 11th International Symposium on Open Collaboration* (15:1–15:4). OpenSym '15. doi:10.1145/2788993.2789846
- Jackson, C. B., Crowston, K., & Østerlund, C. (2018). Did they login?: Patterns of Anonymous Contributions in Online Communities. *Proceedings of the ACM on Human-Computer Interaction (CSCW18)*, 1–16. doi:10.1145/3274346
- Krishnamurthy, S., Ou, S., & Tripathi, A. K. (2014). Acceptance of monetary rewards in open source software development. *Research Policy*, 43(4), 632–644. doi:10.1016/j.respol.2013.10.007
- Lewoniewski, W., Węcel, K., & Abramowicz, W. (2017). Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles. *Informatics*, 4(4), 43. doi:10.3390/informatics4040043
- Marwell, G., & Oliver, P. (1993). *The critical mass in collective action: A micro-social theory*. Cambridge, UK: Cambridge University Press.
- Meng, B., & Wu, F. (2013). COMMONS/COMMODITY: Peer production caught in the Web of the commercial market. *Information, Communication & Society*, 16(1), 125–145. doi:10.1080/1369118X.2012.675347
- Morgan, J. T., & Halfaker, A. (2018). Evaluating the impact of the Wikipedia teahouse on newcomer socialization and retention. In *Proceedings of the 14th International Symposium on Open Collaboration* (20:1–20:7). OpenSym '18. event-place: Paris, France. doi:10.1145/3233391.3233544
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Oreg, S., & Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, 24(5), 2055–2073. doi:10.1016/j.chb.2007.09.007
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.

- Panciera, K., Halfaker, A., & Terveen, L. (2009). Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 51–60). GROUP '09. doi:[10.1145/1531674.1531682](https://doi.org/10.1145/1531674.1531682)
- Piskorski, M. J., & Gorbatâi, A. D. (2017). Testing Coleman's social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology*, *122*(4), 1183–1222. doi:[10.1086/689816](https://doi.org/10.1086/689816)
- Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, *1*(1), 13–32.
- Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, *52*(7), 1000–1014. doi:[10.1287/mnsc.1060.0553](https://doi.org/10.1287/mnsc.1060.0553)
- Thebault-Spieker, J., Hecht, B., & Terveen, L. (2018). Geographic Biases are 'Born, not Made': Exploring Contributors' Spatiotemporal Behavior in OpenStreetMap. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork* (pp. 71–82). doi:[10.1145/3148330.3148350](https://doi.org/10.1145/3148330.3148350)
- Warncke-Wang, M., Ranjan, V., Terveen, L., & Hecht, B. (2015). Misalignment between supply and demand of quality content in peer production communities. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media (ICWSM '15)* (pp. 493–502). AAAI. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10591/10532>
- Willer, R. (2009). Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem. *American Sociological Review*, *74*(1), 23–43. doi:[10.1177/000312240907400102](https://doi.org/10.1177/000312240907400102)
- Wu, C.-G., Gerlach, J. H., & Young, C. E. (2007). An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management*, *44*(3), 253–262. doi:[10.1016/j.im.2006.12.006](https://doi.org/10.1016/j.im.2006.12.006)

## 8. ACKNOWLEDGEMENTS

The author wishes to thank her committee: Adrienne Russell and chair Benjamin Mako Hill, who also served as her advisor for this project. Both are faculty in the Department of Communication at the University of Washington. Morten Warncke-Wang of the Wikimedia Foundation provided both valuable advice on multiple occasions and access to his dataset. Aaron Halfaker of the Wikimedia Foundation provided advice in the effective use of the ORES machine learning system for quality analysis. A pilot version of this project was completed as part of the Advanced Regression course taught by

Jeffrey Arnold in the Center for Statistics in the Social Sciences at University of Washington. Members of the Community Data Science Collective provided advice on two early drafts of this project as part of a workshop, including Jeremy Foote, Floor Fiers, Wm Salt Hale, Sohyeon Hwang, Sejal Khatri, Charles Kiene, Sneha Narayan, Aaron Shaw, and Nathan TeBlunthuis. The author gratefully acknowledges support from the National Science Foundation, awards CNS-1703736 and CNS-1703049, and the Sloan Foundation through the Ford/Sloan Digital Infrastructure Initiative, Sloan Award 2018-11356.