Reciprocity and Donation: How Article Topic, Quality and Dwell Time Predict Banner Donation on Wikipedia

RAFAL KOCIELNIK, Human Centered Design & Engineering, DUB Group, University of Washington, USA
OS KEYES, Human Centered Design & Engineering, University of Washington, USA
JONATHAN T. MORGAN, Wikimedia Foundation, USA
DARIO TARABORELLI, Wikimedia Foundation, USA
DAVID W. MCDONALD, Human Centered Design & Engineering, DUB Group, University of Washington, USA
GARY HSIEH, Human Centered Design & Engineering, DUB Group, University of Washington, USA

Donation-based support for open, peer production projects such as Wikipedia is an important mechanism for preserving their integrity and independence. For this reason understanding donation behavior and incentives is crucial in this context. In this work, using a dataset of aggregated donation information from Wikimedia's 2015 fund-raising campaign, representing nearly 1 million pages from English and French language versions of Wikipedia, we explore the relationship between the properties of contents of a page and the number of donations on this page. Our results suggest the existence of a reciprocity mechanism, meaning that articles that provide more utility value attract a higher rate of donation. We discuss these and other findings focusing on the impact they may have on the design of banner-based fundraising campaigns. Our findings shed more light on the mechanisms that lead people to donate to Wikipedia and the relation between properties of contents and donations.

CCS Concepts: • **Human-centered computing** → Empirical studies in HCI; Empirical studies in collaborative and social computing; Wikis

KEYWORDS

Donations; Fundraising Campaigns; Modeling; Prosocial Computing; Reciprocity; Wikipedia

ACM Reference format:

Rafal Kocielnik, Os Keyes, Jonathan T. Morgan, Dario Taraborelli, David W. McDonald, and Gary Hsieh. 2018. Reciprocity and Donation: How Article Topic, Quality and Dwell Time Predicts Donation on Wikipedia. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 2, CSCW, Article 091 (November 2018). ACM, New York, NY. 19 pages. https://doi.org/10.1145/3274360

1 INTRODUCTION

While advertisement- and subscription-based revenue continue to be the primary ways to monetize websites, many sites also rely on user donations as an important part of their revenue [17,51]. One well-known site that relies on individual donations is the Wikipedia – the free encyclopedia that anyone can edit [48], operated by the non-profit Wikimedia Foundation.

Author's addresses: Rafal Kocielnik (rkoc@uw.edu), Os Keyes (okeyes@uw.edu), David W. McDonald (dwmc@uw.edu) and Gary Hsieh (garyhs@uw.edu), Human Centered Design & Engineering, University of Washington, Seattle, USA; Jonathan T. Morgan (jmorgan@wikimedia.org) and Dario Taraborelli (dario@wikimedia.org), Research, Wikimedia Foundation, San Francisco, California, USA.

 2573-0142/2018/November – ART091 https://doi.org/10.1145/3274360

In the 2016-2017 campaign Wikimedia raised $91 million from 6.1 million donations, an increase from the 2015-2016 campaign where $77 million was raised. The donation-based revenue is especially important and useful for these sites that value content neutrality [2,22]; not relying on advertisement revenue can help preserve their process integrity and independence[1].

One of the primary strategies to solicit donation is through fundraising banners [26]. As stated in their Fundraising Report (2016-2017), banners are an integral part of their fundraising model, which "facilitates the transition from reading Wikipedia, to wanting to contribute to the site through a donation." [55] Further, through their survey studies, they have found that "Wikipedia readers don't consider our fundraising content too intrusive or aggressive." But as with all fundraising campaigns, there continues to be a concern of burnout [28,30]. Through extensive A/B testing of banner content, the fundraising team has been able to improve the overall efficiency of these banners, reducing both the banner impressions during the campaign, as well as the amount of time to reach the fundraising goal.

However, an important part of the banner-based fundraising, that may have been overlooked, is how properties of specific pages on which banners are displayed influence or predict donation behavior. Different pages may draw different visitors for different reasons: the "Kim Kardashian" page on Wikipedia may be attracting a different type of visitor with a different purpose than the "Computer-supported cooperative work" page. Are there potential systematic differences in donation rates across pages? What are some of these differences? Exploring these questions may lead to more effective use of donation banners that are tailored to page contents. This can further improve the efficacy of these fundraising efforts while minimizing burnout.

In this work, we explore how different properties of a Wikipedia page predict donation behaviors. Specifically, using aggregated banner fundraising data from the 2015-2016 campaign for both the English and French language versions, we tested how topic category, page quality and type, and page dwell time affects donation. Our findings suggest a reciprocity hypothesis in donation: that people give because they received or anticipate future benefits [38,40:5]. In this context, the benefits can stem from content that offers task-oriented utility [27]. Congruent with this view, we found that articles of higher quality (e.g., completeness, informativeness, and accuracy [50]), and those that provide task-oriented contents (i.e., articles related to academic or professional activities than to leisure, reading of which has also been found to be more often motivated by work or school tasks [44]) attract a higher rate of donation. Relatedly, we found that pages that ask users for extra interaction steps to reach the contents lead to a much lower donation rates. These are pages such as redirect, list and disambiguation that may be thought of as impeding users' task progress. These have a much more negative impact on donation rates than even the pages with low quality content.

This work's main contributions are the following:

1) Using real-life Wikipedia donation datasets for two language versions we show the existence of a reciprocity mechanism in donation.
2) We demonstrate three measurable indicators useful for predicting donation decisions.
3) We propose ways of redesigning donation banner and donation campaigns to take advantage of our findings.

[1] https://en.wikipedia.org/wiki/Wikipedia:Funding_Wikipedia_through_advertisements

2 RECIPROCITY AND DONATION

The central premise of this work is that during banner-based campaigns, the pages on which the banner resides can attract different people and/or different uses. And due to these differences in people and uses, the rates of contribution on the pages can vary in some systematic way.

There are two recent works that support the first part of the above claim. One is a paper linking people's personal value with topical interests. Specifically, the research found that those holding stronger universalism values are more interested in reading environmental articles, and those with stronger achievement values are more interested in work-related articles [25]. This result suggests that different content attracts different types of individuals. The second paper examines why people read Wikipedia [44]. The findings suggest two primary motivation groups for reading Wikipedia, boredom and satisfying work or school related needs. Specifically, those who use Wikipedia for work and school are more likely to visit pages such as war, history, or mathematics, over pages about sports. This strand of research suggests that different pages on Wikipedia attract different types of people and uses.

What is yet to be studied, and the main focus of this work, is whether these potential systematic differences in users and usages would lead to different donation behaviors on these pages and what page-level features may predict that. Consequently, our main research question is framed as follows: **RQ. Do donations across Wikipedia pages vary in some systematic ways?**

Next we turn to the literature on charitable donations to understand how donation rates might vary across pages. Much prior research has studied why people make charitable contributions [2,22,40]. These works have identified a number of motivations people might have when making a donation, such as: altruism - increasing another's welfare without any external rewards [6]; impure altruism – giving motivated by increasing one's positive emotional feeling or warm-glow [1,2]; peer pressure, authority [46]; prestige, respect, friendship, and other psychological objectives [4,37]; social acclaim or avoidance of scorn [3,35]; image or reputation considerations [42], increase in one's self-esteem [13], as well as income or tax benefits [13,36].

Surveys have also been conducted specifically on Wikipedia for why people do and do not donate to Wikipedia [18] indicating important motivations such as passion for organization's mission or cause. Furthermore, experiments conducted by Wikipedia itself showed that e.g., a message from the site's founder, Jimmy Wales, was more likely to encourage a donation than similar messages not attributed to Wales[2], which relates to the principle of authority mentioned earlier [46]. This body of literature suggests a number of factors that influence or predict people's likelihood to donate. However, many of these factors would not suggest that rates of donation contribution would vary systematically across pages on Wikipedia.

First, some of these factors may not be applicable in the context of Wikipedia banner campaigns. For example, peer pressure or friendship should not be a factor in this context as these donation solicitations are coming from Wikimedia as an organization. Similarly image motivation or reputation considerations as well as avoidance of scorn or desire to receive social acclaim or prestige should not be motivations here, as people's contributions in the banner campaigns are not publicized by default (people can opt to share it with their social network). Even if shared, these image benefits should be the same regardless of what page triggered the donation in the first place.

This brings us to the second consideration, that other factors, though they may influence donations on Wikipedia in general, should not result in different contribution rates across

[2] Information is Beautiful (2010) – https://informationisbeautiful.net/2010/the-science-behind-wikipedias-jimmy-appeal/

pages. Prime examples are impure altruism and self-esteem motivations as there is no reason why a particular page would trigger a stronger positive emotional feeling or warm-glow after donation than any other page. Similarly the appeal of authority, such as a Wikipedia's founder, should have the same effect regardless of the page on which it is presented, as this authority represents and promotes Wikipedia as an entire organization. Same is true when considering a motivation to support Wikipedia's cause or mission statement, both stay the same and there is no reason why certain pages would lead to higher motivation related to organization's cause than others. Finally income or tax benefits are naturally not affected by a particular page on which a donation was given.

On the other hand, one reason for donation that has been widely studied and that could predict differences in donation contributions across pages is *reciprocity*. Reciprocity is a social norm. Reciprocity suggests that people should respond in-kind to others: help others who help them [19]. Extended to the context of charitable giving, people may be both compelled by the reciprocity motive because "they have benefited from the charities' activities in the past or anticipate the need for their services in the future."[13]. Prior work from both controlled experiments and field studies have demonstrated the consistency and strength of reciprocity [10,16,24,29].

Research has found that given a reciprocity motive, the way to increase one's compliance with a request is by increasing their sense of indebtedness. For example, in a controlled study, participants who received a soft drink from a confederate were more likely to comply with their requests [41]. Similar effects have also been extended in field studies of charitable contributions. For example, in a field study soliciting donations, experimenters found that offering a small gift increased frequency of donations by 17 percent, and that offering a big gift increased frequency of donation by a whopping 75 percent [15].

2.1 Hypothesis formulation

Applied in this context, we hypothesize that pages that offer higher utility would attract more donations. There are two possible mechanisms for this. One mechanism, suggested by the Hsieh et al. paper on topical interests, is that people who more frequently use Wikipedia for functional purposes, self-select to visit higher utility pages more frequently [25]. They may generally be more indebted to Wikipedia and thusly feel more motivated to contribute. An alternative mechanism, not due to self-selection, is that people are more likely to contribute when a specific instance of use offers more utility. By increasing utility for visitors on that specific page, potential contributors feel more motivated to contribute (again due to a heightened sense of indebtedness).

Based on this line of reasoning, we posit that pages that provide more utility attract more donations. There are a number of page-level features that might predict such higher utility. The Singer et al. paper found that the two main types of motivation for using Wikipedia are for a specific task (school/work) and due to boredom. The former type of usage, where visitors are using Wikipedia to find information for specific tasks may result in more direct utility. This prior work found that individuals using Wikipedia for task-oriented purposes are more likely to visit certain task-oriented topics, and generally spend more time on individual pages. Thus, we hypothesize that:

H1: Pages on more task-oriented topics attract more donations.

H2: Pages on which users spent more time attract more donations.

Another potential way of considering information utility is by the quality of information. Editors on Wikipedia put a lot of effort into improving the quality of pages. Such quality improvements can be related to better organization of contents, as well as more reliable and complete information. Pages of higher quality are naturally easier and more valuable to interact with. We therefore also expect that the pages of higher quality will lead the users to appreciate and value such pages more and that these users will be more likely to reciprocate for the effort put into the page by the editors:

H3: Pages of higher quality attract more donations.

Table 1. Descriptive statistics summarizing the data for both language versions of the Wikipedia donation datasets provided by the Wikimedia Foundation.

Language version	Number of pages	Page impressions	Page donations
English	830,695	Min: 100 Max: 13,210,276 Mean: 1,578.59 Median: 407	Min: 0 Max: 71,318 Mean: 2.04 Median: 0
French	174,207	Min: 101 Max: 2,608,374 Mean: 569.78 Median: 243	Min: 0 Max: 5,604 Mean: 0.57 Median: 0

3 COLLECTING PAGE-LEVEL FEATURES

In order to test our hypothesis we were given access to a private dataset from the Wikimedia Foundation[3] with page-level aggregated information about the 2015 donation campaign for French and English language versions of Wikipedia (i.e., donation counts per page). The French campaign ran from October 22, 2015 to November 04, 2015, while the English one ran from October 22, 2015 till December 31, 2015. The original data we were provided included: page title (the title of the page in the original language version), English page title, number of impressions (the number of views of the page during the campaign period), and number of donations (the number of donations given for a page). Descriptive statistics are presented in Table 1. For anonymity reasons our datasets did not contain any individual per user donation information, donation amount, or any specific time information about when each donation has been given.

3.1 Scraping the contents of the pages

We scraped the contents of the pages using the Wikipedia API[4]. As the donation campaign happened in the past, it is likely that the content of the pages changed. Therefore we used the API[5] to obtain the latest version from the history just before the end of the fund-raising campaign. We collected the media-wiki contents for each page and stored it in a JSON file with additional metadata such as: *page id*, *date and time of the revision*, *revision id*, *page length*, *comments from latest edit*, *name* and *user id* of the last editor of the page.

[3] Wikipedia Foundation - https://wikimediafoundation.org/wiki/Home
[4] https://www.mediawiki.org/wiki/API:Main_page
[5] https://www.mediawiki.org/wiki/API:Revisions

As a result of this process we managed to successfully scrape past revisions for 172,268 of the 173,207 (98.89%) pages from the French language version and 805,300 of the 830,695 (96.94%) pages from the English language version of Wikipedia. The cases for which revisions could not be obtained were due to permanent deletion or complete renaming of the page, which can happen, e.g., due to copyright issues or major reorganizations of contents[6].

3.2 Scraping the page topic categories

Based on prior work, two dominant methods exist to determine the topics of the pages. One relies on the use of unsupervised NLP technique called topic modeling (e.g., using an algorithm such as Latent Dirichlet Allocation [8]) as used in [44]. This topic modeling approach is a non-deterministic technique and the results obtained are likely different each time the algorithm is run even on the same dataset and with the same set of parameters [5]. The other technique relies on the use of human curated topic category structure maintained by the Wikipedia editors [31], which has been used in [53].

Relying on Wikipedia provided topic categories comes with its own challenges. The categorization is a directed graph that does not necessarily reflect the topic categorization desired. Sheer number of unique categories, which is estimated at 1.6M provides a challenge. Also, due to the complexity of the structure and volunteer model of providing content there are many imperfections related to this categorization (e.g., cycles in category dependencies) [31]. Nevertheless, this approach allows for greater reproducibility.

In this work we scraped the topical categories for the pages by following up the parent categories starting with the categories assigned directly to the page. Wikipedia API provides an ability to obtain parent categories for each category. Our main purpose was to test our hypotheses of reciprocity and therefore we were looking into categorizing the pages into task and non-task oriented (H1). While Wikipedia's categorization does not differentiate between task-oriented vs. non-task-oriented pages recent work that examined the motivations people have for reading Wikipedia articles have found a number of topics that are likely to be differently accessed by users with task and non-task-oriented motivations [44]. Specifically, topics of *war & history, mathematics, technology, biology & chemistry, and literature & arts* were found to be much more frequently accessed by users who use Wikipedia for work or school motivated purposes. Furthermore, these topics while cover wide range of different areas, are more related to academic or professional activities than for leisure. Hence, we refer to these as task-oriented topics. On the other hand, articles on topics such as *Sports, 21st century and TV, Movies, and Novels* have been found to be more frequently accessed by users who described their motivation as bored/random in this prior work. These topics are also more leisure-oriented. We refer to these as non-task oriented topics.

Two things are worth noting about the topic categorization provided by the prior work. First, the task/non-task categorization is based on general trends across a large number of readers. Given that, a particular individual may access an article on topic categorized as non-task oriented for a specific task related purpose (e.g., a journalist writing an article about history of baseball will access *Sports* related article for work). The categorization, however, offers general useful trends that hold across large number of readers. Second, the way prior work determined the topic categorization is based on examining the topics that readers with certain motivations are more likely to access. This means that only a subset of topics has been categorized (i.e., only the topics

[6] https://en.wikipedia.org/wiki/Wikipedia:Deletion_policy

that popped-up as being systematically more frequently accessed by readers with certain motivation have been reported). In our analyses, we assigned the topics not identified as either task or non-task-oriented to represent the baseline "other topics".

Table 2. Mapping of the task/non-task focused topics into Wikipedia provided categories used in our work.

Category	Topics from [44]	Equivalent Wiki categories
Task-oriented topics	War & history	History & Events
	Mathematics	Mathematics & Logic
	Technology	Technology & Applied Sciences
	Biology & Chemistry	Natural & Physical Sciences
	Literature & Arts	Painting, Photography, Sculpture, Drawing, Poetry
Non-task oriented topics	Sports	Sport
	21^{st} century	*No directly equivalent mapping*
	TV & Movies & Novels	Film, Television, Publishing

In this prior work, in order to determine the topics of the pages, the authors relied on LDA based topic modeling. This approach, as discussed earlier is hard to reproduce [5]. To best approximate their findings in our work, we mapped the topics onto the Wikipedia provided categorization (see Table 2). Through our categorization process, we found that 230,511 English version pages (28.62%) and 55,617 French version pages (32.36%) were categorized as task-oriented. Whereas 83,858 of the English version pages (10.41%) and 20,875 of the French version pages (12.15%) were categorized as non-task-oriented.

3.3 Measuring the contents quality on Wikipedia

A number of past research works have tried to address the problem of assessing the contents quality, especially on Wikipedia [23,50,52]. While Wikipedia does have some quality ratings of pages provided by the editors, these are available only for a very small subset of pages. Consequently, past work has examined fully automated approaches based on the use of general text readability and structuring features [50], textual features related to length, structure and style [23], lifecycle based metrics [52] as well as more recent approaches based on end-to-end deep learning models [12]. Among all these, the Objective Revision Evaluation Service (ORES) has gained particular popularity [50]. Thanks to its interpretability, continuous improvements and the fact that it has already been implemented on 5 different Wikipedia language versions[7], it has become a de-facto standard.

ORES is a set of machine learning models designed to evaluate edits to Wikipedia. Specifically it features an edit quality model [49], which can be used to gauge the quality of an article at a particular point in time. Tests by third-party researchers have found it to be highly performant [12]. It has already been relied on in studies of Wikipedia content dynamics to (amongst other things) evaluate the efficacy of an intervention to improve content coverage of women on Wikipedia [20].

ORES outputs a label corresponding to one of a range of quality classes that are also used for manual quality assessment by the editors. These classes can vary between different Wikipedia

[7] https://tools.wmflabs.org/ores-support-checklist/

language versions, but in the case of the English and French Wikipedias, both are almost identical and based on the core categories defined by the original "Assessment" Wikiproject[8].
Furthermore, upon closer examination, we found that the two lowest quality categories, called Start and Stub actually contain a number of subcategories. These have also been identified on Wikipedia official page categorization[9]. Consequently, we split these two main lowest quality categories into a number of subcategories: *redirect, disambiguation (explicit and implicit), category description, list page, may refer to page, other uses page, and other non-article categories.*
In order to identify these subcategories we searched for specific editors keywords in the contents or title of the pages (e.g., for redirect pages, we searched for #REDIRECT in page's contents, while to identify list pages, we searched for "list of" in the page's title).
The quality categories obtained from ORES together with the extracted subcategories are shown in Table 3.

3.4 Obtaining page dwell time

Page dwell time represents "time on page" – how long (in seconds) a reader spent on a particular page before opening a subsequent one. This can be identified by reconstructing each user's browsing sessions, using the timestamp of each page request and an accompanying unique user ID, and then used as a possible predictor of donation likelihood.
To identify the dwell times for each article in our fundraising dataset, we gathered page views from the internal and private database we were given access to by the Wikimedia Foundation. We gathered page views spanning 1 to 31 December 2017 and excluded those where the user was identified as some form of automata or the user's page views did not include at least one article in our dataset. We then reconstructed the user's sessions, using the methodology developed by Halfaker et al. [21], which approximates sessions by looking for a specific gap in time (usually standardized at one hour) between sequential page requests from a single user. From the reconstructed sessions, we identified the dwell time for the pages. This resulted in slightly over 14 million dwell time values, distributed over 796,000 pages.
These dwell times were aggregated per page, producing the median, arithmetic mean and geometric mean dwell time for each page that both appeared in our database and in the page view data for the given time period. In further analysis we used median dwell time, as it was the least affected by outliers and also offered the strongest predictive power among the aggregations we explored.
The use of December *2017* rather than 2015 does raise some issues: after all, the fundraising and dwell time data covers different periods of time, when we could expect different levels of attentiveness to particular pages (the Wikipedia article on Donald Trump, for example, probably drew much less attention three years ago). Unfortunately, the Wikimedia Foundation does not keep logs with enough information to allow us to calculate dwell time from as far back as 2015[10]; unlike fundraising data, browser session data is sanitized after 90 days in a fashion that prevents us from undertaking session reconstruction.[11] As a consequence we are unable to rely on data that would allow a 1:1 comparison. Instead, we picked from the data available, and aimed to control other possible confounds as much as possible. The most concerning of these was seasonality: seasonal changes in how people browse have long been recognized and studied

[8] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_assessment
[9] https://en.wikipedia.org/wiki/Category:Articles_by_quality
[10] https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews
[11] https://meta.wikimedia.org/wiki/Data_retention_guidelines

[14,34], and so being able to control for it is a common part of methodological design when studying Wikipedia or other platforms [11,54]. Using data from December specifically allows us to control for this by ensuring that the two datasets at least mirror each other seasonally, even if they cannot directly match.

Table 3. ORES quality categories as provided for the English and French Wikipedias and the additional non-article categories we extracted.

English quality category (Highest on top)	**Equivalent French category** (Highest on top)	**Reader's Experience based on Wiki**[8] (English version)
Featured Article (FA)	Article de Qualité (ADQ)	Professional, outstanding, and thorough; a definitive source for encyclopedic information.
Good Article (GA)	Bon Article (BA)	Useful to nearly all readers, with no obvious problems; approaching (but not equaling) the quality of a professional encyclopedia.
B	Article Avancé (A)	Readers are not left wanting, although the content may not be complete enough to satisfy a serious student or researcher.
C	Article bien construit (B)	Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study.
Start	Bon début d'article (BD)	Provides some meaningful content, but most readers will need more.
Stub	Ébauche (E)	Provides very little meaningful content; may be little more than a dictionary definition. Readers probably see insufficiently developed features of the topic and may not see how the features of the topic are significant.
Redirect		A page that serves as an intermediary link to refer to the article page. User may end up on a redirect page e.g., in case of misspelling.
Disambiguation (explicit)		A page that offers only a list of links to other pages that are likely related. It is also denoted by a title with "(disambiguation)" appended at the end.
Disambiguation (implicit)		A page that offers contents related to the most likely topic of interest and a number of additional links for similar pages. It does not have explicit "(disambiguation)" in the title.
Category description		Internal Wiki page with description of a topic category useful mostly for editors.
List page		A list of items, e.g., a list of songs in an album.
May refer to page		A page with minimal contents and small "may refer to" link to similar titled pages.
Other uses page		A page with minimal contents and small "other uses" links to related pages.
Other non-article pages		Other, infrequent non-article pages, such as portal, template, image, project.

3.5 Final dataset

Our final dataset contained fewer pages than the original data shared by the Wikimedia Foundation, due to the variety of reasons described in the previous section. The breakdown of the number of pages lost in the collection process is presented in Table 4.

Table 4. Breakdown of pages retrieved and missing by data type for both language versions.

	English dataset			French dataset		
	Missing	Obtained	% Initial	Missing	Obtained	% Initial
Initial pages		830,695	100%		174,207	100%
Revision contents	**19,575**	**805,300**	**96.94%**	**1,336**	**172,268**	**98.89%**
Topic categories	70,881	734,419	88.41%	6,505	165,763	95.16%
Quality rating	1,394	803,907	96.76%	414	171,854	98.65%
Dwell time	92,921	712,379	85.76%	73,950	100,257	57.55%
All information	**97,412**	**711,683**	**85.67%**	**74,078**	**100,129**	**57.38%**

Table 5. Break down of pages in both language versions by ORES quality categories and the additional non-article categories.

Quality + non-article	English dataset		French dataset	
(Highest on top)	Count	% Total	Count	% Total
FA/ADQ	19,080	2.37%	4,167	2.43%
GA/BA	58,084	7.23%	8,212	4.78%
B/A	79,829	9.99%	3,707	2.16%
C/B	301,151	37.46%	52,277	29.84%
Start/BD	228,597	28.38%	71,943	41.86%
Stub/E	57,195	7.10%	20,161	11.73%
Redirect	27,638	3.43%	4,737	2,76%
Disambiguation (explicit)	2,780	0.35%	735	0.43%
Disambiguation (implicit)	3,503	0.43%	891	0.52%
Category description	792	0.10%	2,001	1.16%
List	14,094	1.75%	3,190	1.86%
May refer to	8,407	1.04%	1,544	0.90%
Other uses	5,816	0.72%	1,072	0.62%
Other non-article	609	0.08%	345	0.20%

In table 5, we break down the number of pages that fall under each of the different ORES quality categories for French and English data. We can see that the most frequent quality categories for English version are C and Start. Together, they comprise 65.84% of the pages. The equivalent quality categories for the French version are also the most frequent, together comprising 71.70%. The best quality equivalent categories for both language versions seem to represent a comparable percentage of the data (2.37% for English, and 2.43% for French).
Table 6 contains a break down the number of pages in different topics categories. We can see that relatively similar percentages of pages were mapped to task-oriented topics for both language versions (26.10% for English and 32.36% for French). The same similarity in proportion can be seen for non-task-oriented pages, with 10.39% present in the English version and 12.15% in the French version.

Table 6. Break down of task and non-task category pages for both language versions.

Topic category	English dataset		French dataset	
	Count	% Total	Count	% Total
Task-related topics	**209,934**	**26.10%**	**55,617**	**32.36%**
War & History	55,942	6.95%	12,005	6.99%
Mathematics	15,640	1.94%	5,302	3.09%
Technology	86,209	10.71%	25,993	15.13%
Biology & Chemistry	47,870	5.94%	76,073	8.05%
Literature & Arts	24,850	3.09%	28,420	4.79%
Non-task related topics	**83,671**	**10.39%**	**20,876**	**12.15%**
Sports	29,328	3.64%	7,590	4.42%
TV & Movies & Novels	54,535	6.77%	13,318	7.75%
Other topics	**511,695**	**63.54%**	**95,361**	**55.49%**

4 ANALYSIS PROCEDURES

For the data transformations described in this section we used Python's *Pandas* and *Numpy*[12] libraries. For statistical analyses, we used the *Statsmodels* library together with verification of the models in *R GLM package*[13]. Here we describe a number of analysis choices and variable transformations undertaken following the exploration of the data.

4.1 Numeric variable transformations

A number of our predictor variables were non-normally distributed. For example, the number of impressions has skewness of 615.06 and kurtosis of 460,800.64, page length a skewness of 4.51 and kurtosis of 41.01 and finally median dwell time a skewness of 5.14 and kurtosis of 32.82. Because non-normal distributions and extreme values can cause numerical instability in linear models [45], we log-transformed these predictors.

4.2 Analysis of correlations between the predictors

We explored the correlations between the predictors as these can also cause numerical instabilities in linear models [7]. While these can be addressed using ridge, lasso or elastic net regularizations, the resulting model coefficients are much harder to meaningfully interpret [56]. To examine if strong dependencies between predictors can pose a problem in our analysis we examined correlations among them. We report correlations after log-transformation. The number of donations was positively correlated with number of impressions (r_s=.592), page length (r_s=.244) and median dwell time (r_s=.149). These correlations are not problematic as the number of donations is our outcome variable. The number of impressions was correlated with page length (r_s=.376), but not correlated with median dwell time (r_s=.053). The median dwell time also wasn't correlated with page length (r_s=.100). Given that these correlations are relatively weak, we did not have to remove any of the predictors or apply any regularization.

Treating the quality as interval scale and running a spearman correlation with the task/non-task-oriented topic pages reveals no relationship (r_s=.024) with non-task-oriented topic pages having slightly lower quality (M=2.73, SD=1.08) than task-oriented ones (M=2.77, SD=1.12). The correlation between the quality and the number of impressions revealed a slightly stronger

[12] https://pandas.pydata.org/, http://www.numpy.org/
[13] https://www.statmethods.net/advstats/glm.html

relationship (r_s=.335), which suggests that higher quality pages are viewed more often than lower quality ones. Still, these correlations are not very strong and do not pose numerical stability problems.

Examining the relationship between quality categories and median dwell time, revealed that the median dwell time was the highest for higher quality categories FA (*Mdn*=66.5), GA (*Mdn*=52), B (*Mdn*=67) and much shorter for lower quality ones - C (*Mdn*=58), Start (*Mdn*=47) and Stub (*Mdn*=31). Median dwell time was generally decreasing with decrease in quality (r_s=.110).

Finally examining the relationship between task-orientation and median dwell time revealed that people tend to spend more time on task-oriented topic pages (*Mdn*=68.5) as compared to non-task oriented topic pages (*Mdn*=39). Spearman correlation between these two categories only and median dwell time, indicated slightly positive relationship (r_s=.180). For pages on other topics, the median dwell time was somewhere in between (*Mdn*=51). Finally, we found no relationship between the number of impressions and the task/non-task-oriented topic pages (r_s=-.021), indicating that the pages on task-oriented topics aren't viewed more often than non-task oriented topic pages.

In summary, our exploration of the dependencies among the predictors showed that they are weakly correlated and therefore do not pose problems for further analysis.

4.3 Model selection

The dependent variable, number of donations, had a high percentage of zeros – 54% of the English pages and 72% of the French pages received no donations. Given such prevalence of 0s, the non-normal distribution, and the count nature of our dependent variable, we decided to explore the use of Poisson and Negative Binomial (NB) models instead of a regular OLS regression. In our case, because of over-dispersion (i.e., variance, σ^2=7267.64, is much larger than the mean, μ=2.26), NB is recommended [43]. Indeed, when we used a likelihood ratio test to compare the best fitted models from these two families, we found that the NB model offered a better fit to the data ($\Delta\chi 2$=398,508, df=1, p<.001).

Given these results, we performed all our analysis by fitting variations of a Negative Binomial models to our data. While exploring the variations of the model with different predictors, we used $\chi 2$ likelihood ratio tests to evaluate if added predictors offered a significant improvement in likelihoods given the associated reduction in the degrees of freedom. We tested the new model against a baseline model with only number of visits, page length and language features. The $\chi 2$ statistic itself represents a change in log-likelihood and can be used to compare the models and therefore gauge the overall predictive power of each separate feature group. To also track model absolute fitness, we used a generally accepted McFadden R^2. This statistic will equal zero if all coefficients are zero (intercept model). It will come close to 1 if the model is very good. Menard [39] argues for McFadden R^2 over other Pseudo R^2 measures on the grounds that it is conceptually closest to OLS R^2 i.e., it reflects a proportionate reduction in the quantity actually being minimized, -2*Log-likelihood.

5 RESULTS

To explore the impact of different factors on the number of donations we fitted a number of Negative Binomial models, to test each predictor group separately and then we also combined all the predictors in one Model (Table 7). Our baseline model had only page length, number of visits and language as predictors. Our combined model offered a statistically significant improvement over the baseline model (χ^2=48,673; *df*=21; p<.001) and in absolute terms a 30.5%

improved over an intercept model (McFadden R^2=305). In the combined model, we note that the number of visits was a strong positive predictor of the donation rates. A 10-fold increase in the number of visits increases the likelihood of donation by 2.625 (p<.001). Similarly, language category, with French compared to reference English, was a negative predictor of donation rates. A French page was 0.753 as likely to receive an additional donation as an English page (p<.001). Finally, a 10-fold increase in character count decreased the likelihood of donation by 0.913 (p<.001). In the next subsections we describe the findings for the individual hypotheses.

Testing H1: Pages that provide more task-oriented value attract more donations.

To test this hypothesis, we look at the predictive power of task-related topic features: *War & history, Mathematics, Technology, Biology & Chemistry* (together 265,551 pages) and non-task-oriented topic features*: Sports and TV & Movies & Novels* (together 104,547 pages). The reference comparison, are all the other topic category pages (607,046 pages). We found all the task-oriented topics increasing the odds of receiving a donation and all the non-task-oriented topics decreasing such odds (**H1 supported**).

As hypothesized, task-oriented pages have a significant positive impact on the number of donations. Among these, the pages related to Mathematics have the highest impact, with each such page being 1.679 as likely (p<.001) to receive an additional donation as a reference pages.

In addition, the non-task-oriented pages have a negative impact on the likelihood of a donation. Pages about Sports are only 0.527 times as likely to result in an additional donation as the pages on other topics (p<.001). Similarly, pages about TV & Movies & Novels are 0.648 as likely to result in an additional donation compared to the other pages (p<.001).

Testing H2: Pages on which users spent more time attract more donations.

In the combined model (Table 7), we found the dwell time – the time user spent on a page – to be a significant positive predictor of the number of donations (**H2 Supported**). Each 10-fold increase in the number of seconds spent on a page translates to 1.158 times increase in the likelihood of a donation (p<.001).

Testing H3: Pages with higher quality contents attract more donations.

We found that compared to the baseline highest quality category, all except for the Quality B/A attracted significantly fewer donations. For example, the second highest category GA/BA was a significant negative predictor of donation rates – it was 0.866 as likely to receive a donation as the best quality FA/ADQ page (p<.001). The two lowest quality categories Start/BD and Stub/E offered larger drops of 0.827 and 0.778 respectively (p<.001 for both). Interestingly, however, the fourth best quality category (C/B) offered a slightly smaller drop (0.964, p<0.001) than the second quality category (.866, p<.001). Also, the third best quality category (B/A) did not offer a significant drop against the best quality category 1.014, *p*=n.s. Overall, however none of the lower quality categories offered a significant increase in the likelihood of a donation (**H3 supported)**.

We further explored the non-article categories we extracted from pages originally categorized as Start or Stub by ORES. As expected we can see that almost all of them result in a significant drop in the likelihood of donation as compared to the best quality category. Specifically, the non-article pages with the highest negative impact were, in order: the category description page (.019, p<.001), the redirect page (.202, p<.001), the list page (.543, p<.001) and the explicit disambiguation page (.578, p<.001). These pages resulted in bigger drops than any of the article quality categories. On the other hand, a number of non-article category pages offered a drop in likelihood of a donation much lower than some of the actual article pages. These were in order: May refer to (1.008, *p*=*n.s.*), Other uses (.942, *p*<.001), and Implicit disambiguation (.919, *p*<.001).

Table 7. Predicting the number of donations on the page using topic, quality and dwell time features.

	Negative Binomial Model (N = 811,812 pages) *Exp(B)*
Page length (characters)[1]	.913***
Number of visits[1]	2.625***
Language version (reference: English)	
Language: French	.753***
Median dwell time (seconds)[1]	1.158***
Topic category (reference: Other topics)	
Task: War & history	1.121***
Task: Mathematics	1.679***
Task: Technology	1.469***
Task: Biology & Chemistry	1.307***
Task: Literature & Arts	1.254***
Non-Task: Sports	.557***
Non-Task: TV & Movies & Novels	.648***
Quality categories (reference: highest quality (FA/ADQ))	
Quality (GA/BA)	.866***
Quality (B/A)	1.014
Quality (C/B)	.964***
Quality (Start/BD)	.827***
Quality (Stub/E)	.778***
Redirect	.202***
Disambiguation (explicit)	.578***
Category description	.019***
List page	.543***
Disambiguation (implicit)	.919**
May refer to ...	1.008
Other uses ...	.942**
Other pages	.705***
$\Delta\ \chi^2$ change against baseline (df)	**48,673 (21)*** **
McFadden R^2	**.305**

*[1] log-transformed due to right skewed distribution, *p <0.05, **p<0.01, ***p<0.001*

6 LIMITATIONS

A number of limitations may have affected the presented results. First and foremost, our dataset has been obtained in a non-experimental manner and therefore our findings are correlational. Complexity of human behaviors and decision-making processes, as well as the aggregated nature of our data resulted in a relatively small predictive power of our models. In terms of predictor variables, topic categorization of Wikipedia content is an active research area and our task/non-task-oriented topic categorization is unlikely to be perfect. Finally, we could not obtain dwell time information for 11.65% of English and 41.74% of French language pages (although the effect of other predictor variables did not change much with and without dwell time in our model tests).

7 DISCUSSION

Our work explores the relationship between properties of the Wikipedia pages and the donation rates on those pages. This research complements prior surveys to study donation motivations on Wikipedia [18], as well as Wikimedia Foundation's internal effort to explore the effects of more than 800 different fundraising banner designs through A/B testing [57]. We discover a reciprocity mechanism, which leads users to be more likely to donate on pages that provide more value to them. While prior work on charitable giving suggests a number of general donation mechanisms, reciprocity among them, this mechanism has not been explored in the context of Wikipedia donation and banner campaigns in general. Better understanding of these mechanisms may in practice lead to more effective use of donation banners that can be tailored to pages contents. This can in turn improve efficacy of fundraising campaigns while minimizing burnout. As we will discuss below, our findings not only demonstrate the effect of reciprocity in Wikipedia donations, but also highlight the potentials for leveraging page-level features to predict and even design more effective banners. Our findings do not apply only to Wikipedia, but to any service that wants to provide free and unbiased information online without relying on advertising.

7.1 Reflection on Reciprocity and Wikipedia Donations

Our results show the existence of reciprocity as a factor that impacts donation decisions on Wikipedia. The social norm of reciprocity suggests that people should help others who helped them [2]. This extended to the context of Wikipedia suggests that people will be more likely to donate to pages that offer higher utility. In our work we captured the proxies for page utility in form of task/non-task oriented topics, quality and time spent on a page. We show that donation rates are indeed higher for: 1) task-oriented topics that provide utility value for users, 2) the pages which users spend more time on as the time spent was an indicator of the usefulness of the contents, and 3) contents of higher quality that is more comprehensive and easier to absorb. Here we discuss these mechanisms in more details.

In our full model with all predictor variables, we observed a negative impact of page length on donations. A few explanations are possible. It might be that the longer the page the harder it is to organize the information on it in such a way that all the contents is equally easy to find. Such difficulty in finding content may lower the perceived value of the page and result in lower likelihood of donation. Another possibility is based on an indication from prior work [44], which reports that longer pages are also the popular ones and users are more likely to be already familiar with the content on such pages prior to their visit. Given that users with such prior knowledge do not learn as much new information by visiting the page, they may feel less of a need to compensate Wikipedia for the contents they have read.

Another aspects worth discussing relates to a more complex than expected relation between quality and donations. We expected a simple linear relation in which the higher the quality, the more likely the donation. One factor here is that the ORES score is still a work in progress. The exact differentiation between categories may not be quite clear and likely not linear. There may also be another force, other than reciprocity at play here. It is possible that people may be motivated to donate to the lower quality pages – if there is at least some content so they can envision the potential value of the page. In other words, the lowest quality categories (C/B/BD/E) generally do not have much content and offer little immediate and anticipated value. Whereas some of middle-level quality pages (B/A) may both signal that they should not be compared with the best quality pages (FA) as they are still works in progress, but also have sufficient content to help readers envision what it could become. These higher than expected donation rates for underdeveloped pages could also potentially be explained by an additional mechanisms such as anticipated reciprocity or emergent need for donation [9]. Such pages will encourage the reader to contribute to help in their further development. Of course, additional research is needed to explore this possibility and examine the relationship between quality and donation in more detail.

In our dataset we separated a number of non-article pages, such as redirect, disambiguation, category description and others. Not surprisingly almost all these non-article pages had a negative impact on the donation rates when compared to best quality articles. While this is hardly surprising, we noticed an interesting pattern. The pages with the lowest odds of donation were the ones that introduced an additional interaction step before user could get to the desired contents, e.g., redirect, explicit disambiguation, list, as well as pages that presented unrelated contents – internal topic category description pages. On the other hand, the pages with the smallest drop in donation rates were the ones that offered actual contents and an additional link to obtain further information e.g., "May refer to…", "Other uses…", and "implicit disambiguation" pages. This can further support the reciprocity hypothesis, as the pages that provide contents are likely to offer more practical value to the reader.

One aspect that is unanswered in our work, is whether people are donating on task-oriented pages because these pages attract those that are generally using Wikipedia for task-oriented purposes (generally indebted to Wikipedia), or because task-oriented pages offer more instantaneous value for those who happened to land on these pages. As we will discuss in the practical uses section below, this can have important implications for how one approaches the design. Similarly, especially in the context of topics, our findings do not completely rule out the impact of other factors such as altruism. As prior work has found that people with more altruistic, self-transcendent values, are attracted to different topics [25]. But altruism alone would not be able to explain our findings especially when it comes to observed impact of quality and dwell time.

One way these potential additional mechanisms might be teased apart is if we have data on when the donation contributions are made relative to when the page loads. If the donation occurs when the page first loads, it is unlikely that the specific page influenced donation decision and would suggest the former explanation (self-selection hypothesis), whereas if the donation occurs after "use", it would offer more credence to the latter explanation.

7.2 Potential uses of the findings in redesigning fund-raising banners

While the models we present explain only a fraction of user behavior, given the non-experimental nature of our data this is expected and consistent with similar work on predicting prosocial behaviors among children (reported R^2 of 0.26) [32] It is also common in the field of econometrics, where a large-scale (often involving thousands of people) social behavior patterns and effects of social campaigns are examined and the reported R^2 are at the levels of 0.2-0.3 [47]. Despite these small effect sizes, the practical impact of our findings can be substantial as Wikipedia articles are viewed more than 500 million times every single day[14]. Given such numbers, even small effects we observed could translate to tens of thousands of additional donations per day. We therefore discuss a few possible practical redesigns of Wikipedia donations campaigns informed by out findings.

[14] https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics

7.2.1 Reinforcing indebtedness Our H1 shows that task-oriented pages trigger the feeling of indebtedness that motivates readers to pay back for the value they obtained from reading such page. This indebtedness can be explicitly reinforced for the non-task-oriented pages through value-tailored messaging [33]. Consider the page on Star Wars, which is on topic related to TV & Movies & Novels (non-task oriented). The message on the fund-raising banner could then be tailored to try to highlight its task-oriented value, e.g., *"DEAR WIKIPEDIA READER: Did you know that reading about science-fiction can boost your creativity? This is one of the ways free contents on Wikipedia can be valuable to you. To protect our independence we will never run ads..."*[15].

In a similar fashion, in accordance with findings in our H3, for pages with high quality, the banner message could try to emphasize the effort put by the editors into making the page high quality and how such effort provides more value due to more comprehensive and well organized content.

These design ideas try to introduce the feeling of indebtedness or further reinforce such feeling if it already exists based on the effects found in this work.

7.2.2 Triggering anticipated reciprocity The anticipated reciprocity takes place when the person offers contribution in expectation of future, not current benefits [9]. As suggested by findings from H3, one could use such mechanism on task-oriented, but currently low quality pages, by explicitly linking the donation to the anticipated improvements in the quality of the page such donation would enable. A banner message tying to such motivation could look like this: "*The page you are viewing could become much more comprehensive and valuable to you, but this is only possible with your donation to support the Wikipedia to further grow and improve...".* Similarly, our finding that some non-article facilitation pages (e.g., redirects, explicit disambiguations) currently result in lower donation rates could possibly be addressed with anticipated reciprocity. A banner message on a redirect or explicit disambiguation pages could emphasize that with more donations, improved algorithms can be developed that could help the reader reach the sought-after contents more reliably without having to go through redirect or a disambiguation: "*You see this page because we were not sure which contents exactly you are looking for. Your donation could improve our algorithms and help you reach what you are looking for much faster".* The redesigns try to position the donation as a necessary action to transform the pages of lesser value into more valuable ones.

8 CONCLUSION & FUTURE WORK

In our work, based on analysis of almost 1 million English and French pages from the 2015 Wikipedia donation campaign, we demonstrated that page-level features such as topics, quality and dwell time can be predictive of donations. Our findings suggest a mechanism of reciprocity, which can be leveraged for practical redesign, and more effective use of donation banners. Our finding are valuable not only for Wikipedia, but in fact can easily be generalized to any service that can provide some value for its users, wants to keep information free and unbiased, and wants to be able to operate free of ads. Future work can explore building better algorithms for boosting donation prediction accuracy and dive deeper into more nuanced aspects of the reciprocity mechanism we identified.

[15] We adapted an actual message from a Wikipedia fund-raising banner:
https://meta.wikimedia.org/wiki/Fundraising/2013-14_Report

ACKNOWLEDGMENTS

This work was in part supported by National Science Foundation grant #1348543. We would also like to thank Ellery Wulczyn for his initial support of the work.

REFERENCES

[1] James Andreoni. 1989. Giving with impure altruism: Applications to charity and Ricardian equivalence. *J. Polit. Econ.* 97, 6 (1989), 1447–1458.

[2] James Andreoni. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *Econ. J.* 100, 401 (1990), 464–477.

[3] Dan Ariely, Anat Bracha, and Stephan Meier. 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *Am. Econ. Rev.* 99, 1 (2009), 544–55.

[4] Omar Azfar. 2001. The logic of collective action. In *The Elgar companion to public choice*. Edward Elgar Publishing.

[5] Arun Balagopalan. 2012. *Improving topic reproducibility in topic models*. University of California, Irvine.

[6] C. Daniel Batson. 2012. 12 A history of prosocial behavior research. *Handb. Hist. Soc. Psychol.* (2012), 243.

[7] David A. Belsley. 1991. *Conditioning diagnostics: Collinearity and weak data in regression*. Wiley.

[8] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, Jan (2003), 993–1022.

[9] Gee-Woo Bock, Robert W. Zmud, Young-Gul Kim, and Jae-Nam Lee. 2005. Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Q.* (2005), 87–111.

[10] Robert B. Cialdini. 2009. *Influence: Science and practice*. Pearson education Boston, MA.

[11] Giovanni Luca Ciampaglia and Dario Taraborelli. 2015. MoodBar: Increasing new user retention in Wikipedia through lightweight socialization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 734–742.

[12] Quang-Vinh Dang and Claudia-Lavinia Ignat. 2017. An end-to-end learning solution for assessing the quality of Wikipedia articles. In *Proceedings of the 13th International Symposium on Open Collaboration*, 4.

[13] Scott Dawson. 1988. Four motivations for charitable giving: Implications for ma. *Mark. Health Serv.* 8, 2 (1988), 31.

[14] Fabon Dzogang, Thomas Lansdall-Welfare, and Nello Cristianini. 2016. Seasonal Fluctuations in Collective Mood Revealed by Wikipedia Searches and Twitter Posts. In *ICDM Workshops*, 931–937.

[15] Armin Falk. 2004. Charitable Giving as a Gift Exchange-Evidence from a Field Experiment. (2004).

[16] Ernst Fehr and Simon Gächter. 2000. Fairness and retaliation: The economics of reciprocity. *J. Econ. Perspect.* 14, 3 (2000), 159–181.

[17] Niall Firth. 2012. *Crowdfunding successes show value of small donations*. Elsevier.

[18] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. 2010. Wikipedia survey–overview of results. *U. N. Univ. Collab. Creat. Group* (2010).

[19] Alvin W. Gouldner. 1960. The norm of reciprocity: A preliminary statement. *Am. Sociol. Rev.* (1960), 161–178.

[20] Aaron Halfaker. 2017. Interpolating Quality Dynamics in Wikipedia and Demonstrating the Keilana Effect. In *Proceedings of the 13th International Symposium on Open Collaboration* (OpenSym '17), 19:1–19:9. DOI:https://doi.org/10.1145/3125433.3125475

[21] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *Am. Behav. Sci.* 57, 5 (2013), 664–688.

[22] William T. Harbaugh. 1998. What do donations buy?: A model of philanthropy based on prestige and warm glow. *J. Public Econ.* 67, 2 (1998), 269–284.

[23] Daniel Hasan Dalip, Marcos André Gonçalves, Marco Cristo, and Pável Calado. 2009. Automatic quality assessment of content created collaboratively by web communities: a case study of wikipedia. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, 295–304.

[24] Elizabeth Hoffman, Kevin A. McCabe, and Vernon L. Smith. 1996. On expectations and the monetary stakes in ultimatum games. *Int. J. Game Theory* 25, 3 (1996), 289–301.

[25] Gary Hsieh, Jilin Chen, Jalal Mahmud, and Jeffrey Nichols. 2014. You read what you value. In *Conference on Human Factors in Computing Systems - Proceedings.* DOI:https://doi.org/10.1145/2556288.2557201
[26] Gary Hsieh and Rafal Kocielnik. 2016. You get who you pay for: The impact of incentives on participation bias. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 823–835. DOI:https://doi.org/http://dx.doi.org/10.1145/2818048.2819936
[27] Chris S. Hulleman, Olga Godes, Bryan L. Hendricks, and Judith M. Harackiewicz. 2010. Enhancing interest and performance with a utility value intervention. *J. Educ. Psychol.* 102, 4 (2010), 880.
[28] Tamieka Jameson. 2017. Communication strategies as drivers of nonprofit donor retention. Walden University.
[29] John H. Kagel and Alvin E. Roth. 2016. *The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics.* Princeton university press.
[30] Beth Kanter and Aliza Sherman. 2016. *The happy, healthy nonprofit: Strategies for impact without burnout.* John Wiley & Sons.
[31] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1509–1512.
[32] George P. Knight, Lora G. Johnson, Gustavo Carlo, and Nancy Eisenberg. 1994. A multiplicative model of the dispositional antecedents of a prosocial behavior: Predicting more of the people more of the time. *J. Pers. Soc. Psychol.* 66, 1 (1994), 178.
[33] Rafal Kocielnik and Gary Hsieh. 2017. Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers. In *CSCW*, 2193–2207. DOI:https://doi.org/http://dx.doi.org/10.1145/2998181.2998324
[34] Tomáš Kramár and Mária Bieliková. 2014. Context of seasonality in web search. In *European Conference on Information Retrieval*, 644–649.
[35] Nicola Lacetera and Mario Macis. 2010. Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *J. Econ. Behav. Organ.* 76, 2 (2010), 225–237.
[36] David Leonhardt. 2008. What makes people give. *N. Y. Times* 9, (2008).
[37] Jonathan Meer. 2011. Brother, can you spare a dime? Peer pressure in charitable solicitation. *J. Public Econ.* 95, 7–8 (2011), 926–941.
[38] Stephan Meier. 2006. A Survey of Economic Theories and Field Evidence on Pro-Social Behavior. *SSRN Electron. J.* (2006). DOI:https://doi.org/10.2139/ssrn.917187
[39] Scott Menard. 2002. *Applied logistic regression analysis.* Sage.
[40] Theodore Millon, Melvin J Lerner, and Irving B Weiner. VOLUME 5 PERSONALITY AND SOCIAL PSYCHOLOGY. 691.
[41] Dennis T. Regan. 1971. Effects of a favor and liking on compliance. *J. Exp. Soc. Psychol.* 7, 6 (1971), 627–639.
[42] David Reinstein and Gerhard Riener. 2012. Reputation and influence in charitable giving: An experiment. *Theory Decis.* 72, 2 (2012), 221–243.
[43] German Rodrıguez. Models for Count Data With Overdispersion. 7.
[44] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. 2017. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, 1591–1600.
[45] Strakoš Z. and Liesen J. 2005. On numerical stability in large scale linear algebraic computations. *ZAMM - J. Appl. Math. Mech. Z. Für Angew. Math. Mech.* 85, 5 (April 2005), 307–325. DOI:https://doi.org/10.1002/zamm.200410185
[46] Behavioural Insights Team. 2013. Applying behavioural insights to charitable giving. *Cabinet Off.* (2013).
[47] Ott Toomet, Siiri Silm, Erki Saluveer, Rein Ahas, and Tiit Tammaru. 2015. Where do ethno-linguistic groups meet? How copresence during free-time is related to copresence at home and at work. *PloS One* 10, 5 (2015), e0126093.
[48] Jakob Voss. 2005. Measuring wikipedia. (2005).
[49] Morten Warncke-Wang, Vladislav R. Ayukaev, Brent Hecht, and Loren G. Terveen. 2015. The Success and Failure of Quality Improvement Projects in Peer Production Communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 743–756. DOI:https://doi.org/10.1145/2675133.2675241

[50] Morten Warncke-Wang, Dan Cosley, and John Riedl. 2013. Tell me more: an actionable quality model for Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*, 8.
[51] Rick Wash. 2013. The Value of Completing Crowdfunding Projects. *ICWSM* 13, (2013), 7th.
[52] Thomas Wöhner and Ralf Peters. 2009. Assessing the quality of Wikipedia articles with lifecycle based metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, 16.
[53] Bowen Yu, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017. Predicting Member Productivity and Withdrawal from Pre-Joining Attachments in Online Production Groups. 1775–1784. DOI:https://doi.org/10.1145/2998181.2998227
[54] Xiaoquan Zhang and Feng Zhu. 2006. Intrinsic motivation of open content contributors: The case of Wikipedia. In *Workshop on Information Systems and Economics*, 4.
[55] 2016-2017 Fundraising Report - Wikimedia Foundation. Retrieved April 19, 2018 from https://wikimediafoundation.org/wiki/2016-2017_Fundraising_Report
[56] How does one interpret coefficients from regularized linear regression, considering multicollinearity present between independent variables? - Quora. Retrieved April 19, 2018 from https://www.quora.com/How-does-one-interpret-coefficients-from-regularized-linear-regression-considering-multicollinearity-present-between-independent-variables
[57] Fundraising/2013-14 Report - Meta. Retrieved April 19, 2018 from https://meta.wikimedia.org/wiki/Fundraising/2013-14_Report

Received April 2018; revised July 2018; accepted September 2018