Public Good Superstars: A Lab-in-the-Field Study of Wikipedia

Center for Law & Economics Working Paper Series

Number 05/2024

Public Good Superstars: A Lab-in-the-Field Study of Wikipedia

Jérôme Hergueux
Yann Algan
Yochai Benkler
Mayo Fuster-Morell

This version: August 2024

Note: This paper has been conditionally accepted for publication in *The Economic Journal*.

To cite this paper:

Hergueux, Jerome, Yann Algan, Yochai Benkler and Mayo Fuster-Morell (forthcoming). Public Good Superstars: A Lab-in-the-Field Study of Wikipedia. *The Economic Journal*.

Public Good Superstars: A Lab-in-the-Field Study of Wikipedia[*]

Jerome Hergueux[†] Yann Algan[‡] Yochai Benkler[§] Mayo Fuster-Morell[¶]

August 28, 2024

Abstract

Many field public goods are provided by a small number of contributors: the "superstars" of their respective communities. This paper focuses on Wikipedia, one of the largest online volunteering platforms. Over 9 consecutive years, we study the relationship between social preferences – reciprocity, altruism, and social image – and field cooperation. Wikipedia editors are quite prosocial on average, and superstars even more so. But while reciprocal and social image preferences strongly relate to contribution quantity among casual editors, only social image concerns continue to predict differences in contribution levels between superstars. In addition, we find that social image driven editors – both casual and superstars – contribute *lower* quality content on average. Evidence points to a perverse social incentive effect, as quantity is more readily observable than quality on Wikipedia.

Keywords: Lab-in-the-Field Experiment; Public Goods; Cooperation; Reciprocity; Social Image; Wikipedia.

JEL Classification: H41, C93, D01, Z13.

[*]This research was funded through the Consolidator grant agreement 647870 from the European Research Council (YA) and the French "excellence initiative" Idex grant agreement W18RAT88 (JH). We thank the Sciences Po médialab, the Wikimedia Foundation and ETH Zurich for logistical support, and Anne l'Hôte, Romain Guillebert, David Laniado, Tarun Chadha and Kene Boun My for excellent research assistance. We are grateful to Dario Taraborelli for his instrumental help in moving this project forward, and indebted to Benjamin Mako Hill and Aaron Shaw for generously sharing some of their data. Our last thank goes to the Wikipedia editors who took part in this study.

[†]*Main & corresponding author.* French National Center For Scientific Research (CNRS, BETA lab), Berkman Klein Center for Internet & Society at Harvard University, and Center for Law & Economics at ETH Zurich. e-mail: `jerome.hergueux@gess.ethz.ch`

[‡]HEC Paris, Department of Economics and Decision Sciences. e-mail: `algan@hec.fr`

[§]Harvard Law School and Berkman Klein Center for Internet & Society at Harvard University. e-mail: `ybenkler@law.harvard.edu`

[¶]Open University of Catalonia and Berkman Klein Center for Internet & Society at Harvard University. e-mail: `mfuster@uoc.eu`

"The problem with Wikipedia is that it only works in practice. In theory, it can never work."
Kizor, Wikipedia administrator.

1 Introduction

From volunteering time in a local charitable organization to running an amateur sports club, and from staffing polling stations on election days to doing regular beach cleanups, many field public goods in our communities are provided by a minority of individuals who exhibit extraordinary levels of dedication to their cause. This fact is perhaps most apparent in online volunteering spaces, where it has been long recognized as a feature of individual involvement with, e.g., Open Source Software (Crowston et al., 2008), Stack Overflow (Wu et al., 2009) or Wikipedia (Viégas et al., 2007).

The skewed structure of participation in many public goods provision contexts gives rise to the emergence of "superstar contributors": highly regarded community members with impressive contribution records. In practice, many volunteering communities are thus articulated around two levels of participation (Shaw and Hill, 2014; Safadi et al., 2021). On the one hand, a minority of "core" contributors display extreme levels of involvement in the project, and play an instrumental role in the functioning of the community. On the other, a large majority of "peripheral" contributors make more discrete contributions, which can nonetheless add up.

Over the past decades, one important stream of the literature on public goods provision has demonstrated the centrality of reciprocal preferences for sustained cooperation. This is true in theory (Rabin, 1993; Fehr and Schmidt, 1999; Sobel, 2005; Falk and Fischbacher, 2006), in the laboratory (Fischbacher and Gächter, 2010; Chaudhuri, 2011), and in the field (Barr and Serneels, 2009; Rustagi et al., 2010; Carpenter and Seki, 2011; Fehr and Leibbrandt, 2011; Leibbrandt, 2012; Gneezy et al., 2016). While, in social dilemma situations, some individuals exhibit purely selfish or (more rarely) altruistic behavior, most endorse the norm of reciprocity: they are willing to cooperate, as long as others respond in kind. This literature therefore emphasizes that the key to sustained voluntary cooperation lies in matching mechanisms that enable strong reciprocity at the group level through, e.g., voluntary association (Page et al., 2005; Cinyabuguma et al., 2005; Charness and Yang, 2014), fiat (Gächter and Thöni, 2005; Burlando and Guala, 2005) or contract design (Kosfeld and Von Siemens, 2011; Bartling et al., 2012).

While powerful, this account of field cooperation based on reciprocal preferences cannot easily explain the emergence of superstars in volunteering communities. The reason is straightforward: strong reciprocity dictates that people try to match the observed contributions of their peers. This can lead to high collective contribution levels in equilibrium, but provides no individual incentives to behave as an outlier by exhibiting the extreme levels of dedication that characterize public good superstars in the field. In theory, the literature puts forward two other social preferences that seem more consistent with such behavior: (i) altruism (Andreoni, 1989, 1990), where one derives utility from providing the public good in and of itself, and (ii) social image motives (Holländer, 1990; Bénabou and Tirole, 2003, 2006; Ellingsen and Johannesson, 2008, 2011), where one derives utility from being able to signal their quality to a large audience, for instance by achieving and striving to maintain a central position in the community of contributors (van Leeuwen et al., 2020; Hong et al., 2021).

The study of the implications of social image concerns for public goods provision, in particular, constitutes another significant stream of the literature on voluntary cooperation, although less developed than that on reciprocal preferences.[1] Becker (1974) long remarked that apparent "charitable" behavior may in fact be motivated by a desire to receive social recognition. Similarly, Bénabou and Tirole (2006) and Olson (2009) have argued that people are often motivated by a desire to win prestige or respect. Empirically, Bursztyn and Jensen (2017) reviewed significant evidence showing that peer signaling can be a strong motivational force shaping various field behavior. Applied to social dilemma situations, Ariely et al. (2009) showed in a real-effort experiment that social image concerns lead to increased contribution levels. Most relevant to our study, van Leeuwen et al. (2020) developed and provided a laboratory test of a model of public goods provision where agents compete to attract social status rents. They showed both theoretically and experimentally that a strong taste for social image leads to the emergence of superstars willing to sustain extraordinary levels of contribution in order to maintain their centrality in the network of contributors.

In this paper, we combine lab and field data from Wikipedia over nine consecutive years to

[1] For instance, in a recent survey of the literature on the "fundamental characteristics and economic consequences of social preferences", Fehr and Charness (2023) dedicate 12 pages to discussing reciprocal preferences, as opposed to two pages for self and social image concerns.

study the motivational correlates of field cooperation in one of the Internet's largest public good provision community. Specifically, we explore how reciprocal, altruistic, and social image preferences account for the dual structure of participation in Wikipedia, i.e., the contributing behavior of "core" versus "peripheral" editors. Over the past 20 years, Wikipedia has emerged as one of humanity's most valuable digital public goods. With 55 million freely usable articles in hundreds of languages (representing about 2,800 printed volumes for the English version alone) and over 20 billion page views a month, its revealed informational value is simply enormous to society.[2] Wikipedia is also a particularly clean environment to study voluntary public goods provision, as (i) individuals self-assign and collectively monitor their work in the absence of monetary incentives (Benkler, 2008; Benkler et al., 2015), and (ii) contributions carry little signaling value on the labor market.[3]

On December 8th, 2011, we ran a lab-in-the-field experiment on the Wikipedia website and recruited a sample of 730 Wikipedia editors to play an online version of the conditional public goods game (Fischbacher et al., 2001). We used this experimental setup to recover laboratory measures of reciprocal and altruistic preferences from our subjects (together with some standard socio-demographic information). In addition, we retrieved the personal user (or "biography") pages of these subjects, and ran a distributed online rating task aimed at measuring their taste for social "signaling" or image.

In a spirit similar to that of Malmendier and Tate (2009) (who identified superstar CEOs in the U.S. based on the awards they received from various media organizations) we relied on Wikipedia's main social recognition device – community awards called Barnstars – to identify the superstar contributors among our subjects (Kriplean et al., 2008; McDonald et al., 2011; Sajnani et al., 2011). Rather than having to (arbitrarily) define the inclusion criteria for this population, our strategy therefore relies on the social recognition practices of the community itself to point us at their most notable members. Finally, within this group of award receiving contributors (i.e., the "superstars") we exploited the fact that some choose to prominently display their Barnstars to their fellow editors to derive a field-based measure of taste for social image, which

[2]For more evidence on the economic value of Wikipedia as well as on the reliability of its encyclopedic content, see https://www.economist.com/international/2021/01/09/wikipedia-is-20-and-its-reputation-has-never-been-higher

[3]Unlike Open Source Software, where labor market signaling has long been identified as one important motive behind individual code contributions (Lerner and Tirole, 2002), no evidence points at Wikipedia contributors using their edits for similar purposes (by, e.g., reporting them on their CV).

complements the one we derived from our distributed user page rating task.

In order to explain field cooperation based on the above covariates, we collected the data publicly available from Wikipedia on the editing activity of our subjects from the time of the experiment, in December 2011, and until November 1st, 2020. Importantly, our study embraces the fact that, in the field, cooperation is typically a multi-dimensional construct. Depending on the social incentives at play, this may lead to conflicting goals at the individual level. For instance, in the workplace, social image motivated workers might invest in signaling their dedication by doing a lot of overtime work – a very visible cooperative activity – at the expense of productivity, which is more difficult to observe. The extant lab-in-the-field literature on public goods provision has long been aware of this dimensionality issue. For instance, in the field context of shrimp catching communities in Brazil, Fehr and Leibbrandt (2011) chose to focus on the quality of individual cooperation by measuring the mesh size of shrimp catchers' nets, as opposed to looking at the mere quantity of resources they extract from the common pool.

In this paper, we take this line of inquiry one step further by asking whether, similar to monetary incentives, social motives can induce substitution (or "perverse incentive") effects between potentially conflicting dimensions of field cooperation. To do so, we rely on the richness of the activity traces available from Wikipedia to study three separate but complementary dimensions of field cooperation: (i) the *quantity* of contributions made, (ii) the *quality* of these contributions, and (iii) the *level of cooperativeness* exhibited towards others while editing.

The main result of our analysis is threefold. First, we find that superstar contributors do appear, on average, more prosocial than casual editors (who are themselves significantly more prosocial than a standard students subjects pool). However, while reciprocal and social image preferences are strongly associated with the quantity of field contributions made among casual editors, only social image concerns continue to relate to differences in contribution levels among the Wikipedia superstars. Furthermore, even within the population of casual contributors, the strength of the relationship between social image concerns and the quantity of field contributions is an order of magnitude higher than that for reciprocal preferences. While our coefficients on altruism are less precisely estimated, we find suggestive evidence that, among superstar contributors, this preference relates to increased levels of cooperativeness towards other editors (i.e., while editing). In a nutshell, our field analysis suggests that reciprocity and social image motives may drive baseline cooperation among the laymen editors who make discrete contributions to the public good, but that only social image concerns may drive the intensive margin of contributions among Wikipedia's superstars.

Second, a comparison of the quantity and the quality dimensions of cooperation in Wikipedia reveals that (i) the correlation between both measures is *negative* at the individual level, and (ii) while they contribute significantly more content, both image-driven superstars *and* regular editors actually produce *lower* quality material on average. We present evidence consistent with the idea that those contributors are likely subject to a perverse incentive effect. Unlike quantity, the quality dimension of editors' contributions is difficult to assess for an external observer. This can be seen empirically: while the total number of community awards received by our subjects strongly relates to the quantity of contributions made as well as to their level of interpersonal cooperativeness, it is not related to their quality.[4] As a result, contributors motivated to earn social recognition substitute quality for quantity in their contributions decisions. These results suggest that, similar to monetary incentives, social incentives can have unintended consequences in terms of how workers arbitrate between possibly conflicting dimensions of cooperation at the group level. These social incentives should therefore be designed with care, e.g., by ensuring that all the relevant dimensions of cooperation in a given organizational context are made equally visible to the community of contributors.

Third, a dynamic analysis of the evolution of our estimates reveals that the relationships between social preferences and field cooperation which we uncover are remarkably stable over time, as most remain statistically significant over the 9 consecutive years covered by our study.

The rest of this paper proceeds as follows.[5] We present our data and variables in Section 2. We conduct a detailed descriptive analysis of this data in Section 3. Section 4 presents our main regression results. Section 5 disaggregates our coefficients to conduct a dynamic analysis of the magnitude of the relationships we uncover over the 9 consecutive years covered by our study. Section 6 concludes.

[4]This is not to say that individual editors do not value the quality of their own work, as can be seen from the fact that our edit quality variable most strongly relates to the length of subjects' user (or "biography") pages on Wikipedia.

[5]In Appendix A, we summarize the lab-in-the-field and Wikipedia literature related to our paper.

2 Data and Variables

We begin this section by describing how we experimentally elicited reciprocal and altruistic preferences within our sample of subjects (Section 2.1). Second, we describe how we relied on the specificity of Wikipedia's social rewarding context to identify the superstar editors in our sample (Section 2.2). Third, we explain how we elicited subjects' social image motives through (i) a distributed social signaling rating task (Section 2.3.1), and (ii) a measure of social image concerns based on field data (only available for superstar editors, Section 2.3.2). Last, we detail the construction of our field dependent variables aimed at capturing (i) the quantity of field contributions made by our subjects (Section 2.4.1), (ii) their level of cooperativeness when editing (Section 2.4.2), and (iii) the quality of their edits (as judged by their peers, Section 2.4.3).

2.1 Eliciting reciprocity and altruism: the online public goods experiment

With support from the Wikimedia Foundation, we used a Wikipedia banner as the recruitment device for our experiment. The Wikimedia Foundation relies on this banner system to advertize its annual fundraising, which makes it relatively familiar even to non Wikipedia contributors. Banners are also used extensively by the community of editors for purposes of internal communication (e.g., to advertise events and other community initiatives). As a result, the banner system is certainly the most powerful and trusted way of reaching out to a wide and diverse audience within Wikipedia. In coordination with the Wikimedia Foundation staff, we coded this recruitment banner so that it would be displayed at the top of every Wikipedia page for logged-in users, until he or she decided either to click on it, or to disable it. (See Figure 1, which features the recruitment banner.)

[FIGURE 1 ABOUT HERE]

Upon clicking on the banner, eligible users were uniquely and automatically identified by the system through their Wikipedia user id, and redirected to the welcome screen of our experimental platform. Users who merely wanted further information about the study could also click a "learn more" button located at the bottom-right of the banner.[6] The experiment was launched on December 8th, 2011 and the banner recruited 730 Wikipedia contributors in 8 hours.[7] Right after the experiment and before payment, we asked subjects for some standard demographic information, i.e., their age, gender, education level and salary range, together with an experimentally validated question on risk aversion (Dohmen et al., 2011).

Our experimental design strictly followed the Internet-specific procedures detailed in Hergueux and Jacquemet (2015).[8] The key experimental game we used to elicit reciprocal and altruistic preferences is the one-shot public goods game. This game is played in groups of four players, each with an initial endowment of \$10. Group members need to decide how much to contribute to a common project. Each dollar invested in the common project produces \$1.6, which is then equally distributed among group members. Thus, a \$1 investment only yields a private return of \$0.4, but benefits all other members of the group. This design captures the social dilemma faced by Wikipedia editors in the field: contributing information to Wikipedia can be individually costly, but is socially efficient.

Following the example of Fischbacher et al. (2001), we elicited two types of contribution decisions: first an unconditional contribution, and then a conditional contribution. For the unconditional contribution, each subject had to decide on his or her contribution in the game described above. For the conditional contribution, each subject determined his or her intended contribution for each possible value (0,1,2, . . . 10) of the average contribution of the three other members of the group. The conditional contributions allowed us to measure the subjects' willingness to behave reciprocally (i.e., to be conditionally cooperative). This design is incentive-compatible since, after the match with other participants has been carried out, one randomly selected decision (i.e., unconditional or conditional) is used to compute the subjects' earnings. The screen eliciting conditional contributions is presented in Figure 2.

[FIGURE 2 ABOUT HERE]

[6] This button redirected to a Wikimedia page where the project was purposefully described in very general, but accurate terms: https://meta.wikimedia.org/wiki/Research:Dynamics of Online Interactions and Behavior.

[7] Through the same protocol, we also recruited a sample of 120 Wikipedia administrators who played a standard Trust game. We report the results of this experiment in Hergueux et al. (2021).

[8] See Appendix B for a detailed account of our experimental procedures.

From there, we classify subjects into four exclusive cooperative types depending on their revealed preferences in the conditional public goods game. To do so, we compute (i) the slope of subjects' reaction functions to the possible average contributions of the other group members (i.e. reciprocity r) and (ii) the average proportion of the endowment that is conditionally contributed across all 11 conditional contributions decisions (i.e. mean contribution m). Our classification rule automates that of Fischbacher et al. (2001) and is consistent with that of Fallucchi et al. (2019), who use hierarchical cluster analysis to separate subjects into exclusive behavioral types based on a meta-analysis of 6 seminal public goods experiments on conditional cooperation:

- *Altruists* contribute their entire endowment, irrespective of the average contribution level of the other players: $m_i = 10$;
- *Reciprocators* match the contribution level of the other players: $r_i \geq 1$;
- *Weak reciprocators* under-match the contribution level of the other players: $r_i < 1$;
- *Free-riders* do not contribute to the public good, irrespective of the average contribution level of the other players: $m_i = 0$.

2.2 Identifying the Wikipedia superstars

To identify superstar contributors within the community of Wikipedia editors, we rely on its main social recognition practice: the Barnstars system.[9] Similar to, e.g., the Congressional Gold Medal in the United States and other similar distinctions of significant symbolic value, a Barnstar is a highly valued symbolic award for a Wikipedia editor. It is typically constituted of an image and title which refer to the type of contribution being acknowledged (e.g., fighting vandalism, mediating conflicts between editors, making an important set of contributions on a given topic etc.) accompanied by a personalized message that precisely acknowledges the set of contributions made to the project by the recipient (see Figure 3 for an example).

[FIGURE 3 ABOUT HERE]

[9] See https://en.wikipedia.org/wiki/Wikipedia:Barnstars

In a seminal paper, Kriplean et al. (2008) carefully described the practice of Barnstar awarding in the community of Wikipedia contributors. Technically speaking, anyone can give or receive a Barnstar. However, the authors showed that, because Barnstars precisely describe the type of contribution being acknowledged, which requires both *hindsight* and *insider knowledge*, this social rewarding practice remains largely limited to experienced editors. Through a significant human coding effort of 2400 randomly selected Barnstars, their paper then describes the range of contributing work most socially valued in the community. General editing activity represents the most acknowledged way of contributing (28%) followed by community support such as leadership, commitment or mentoring (25%), and vandal fighting (11%). Follow-on papers in computer and information science have largely built upon this work by deploying machine learning techniques to further extend and confirm those results (e.g., McDonald et al. (2011) Sajnani et al. (2011)).

Studies in economics have relied on this literature to design image-driven field interventions that randomly awarded Barnstars within specific subsets of the population of experienced (Restivo and Van De Rijt, 2012; Restivo and van de Rijt, 2014), and novice contributors (Gallus, 2017) to study their effect on subsequent editing (see Appendix A.2). By contrast, our design does not manipulate Barnstars. Rather, it exploits this social recognition practice to identify the Wikipedia superstars within our sample of subjects. This strategy allows us to encompass the wide range of contributing activities most acknowledged in Wikipedia in our definition. It also discharges us from having to (arbitrarily) operationalize and balance the many ways in which one may be considered to have reached "superstar" status in this community. Rather, we let the contributors themselves point us at their most prolific and remarkable peers. As a result, within our sample of subjects, we define as "superstars" those who were awarded a Barnstar by another editor over the course of their editing career.[10]

[10]This means that we also collected the Barnstars received by our subjects *after* the time at which we ran our experiment, on December 8th, 2011. This implementation choice is different than the strategy we adopted with respect to our other independent variables of interest, which were measured at the beginning of our time period. As it turns out, in our data, 73 subjects received their first community award after this date. We thus follow the extant Wikipedia literature (Bryant et al., 2005; Panciera et al., 2009), which found that superstar editors start to behave as such from the very beginning of their participation, by classifying these subjects as superstars by default. In Appendix E.4, we further show empirically that, alternatively, excluding these subjects from our sample of superstars leaves our results unchanged.

2.3 Eliciting social image motives

Surprisingly, the extant literature has not yet converged on a standard set of experimental tools to identify social image motives in the laboratory (see, e.g., Henry and Sonntag (2019) for a recent discussion and design proposal.) In this paper, we therefore develop two alternative measures to reveal the heterogeneity of subjects' taste for social image or "signaling": one based on distributed human rating of their personal user (or "biography") pages, and the other based on observed field behavior (only available for the Wikipedia superstars).

2.3.1 The social signaling rating task

All registered users on Wikipedia have the possibility to create a personal user page where they can present themselves to the community, post some general information about their interests, list the articles they helped improve and the like.[11] We rely on this wealth of self-generated "biographical" content to create a subject-specific measure of taste for social image or "signaling" within the Wikipedia community. To do so, we first retrieved the user page of each subject in our sample. Out of our 730 subjects, 3 user pages could not be retrieved because a Wikipedia administrator had deleted their history logs from the database (this can happen, e.g., when users post illicit content to their page, or simply because they wish to permanently erase its content). Of the remaining 727 subjects, 171 never edited their user page, and 12 created a blank one (by, e.g., posting a blank space). This left us with 544 user pages featuring substantive personal content to analyze.

From there, we designed a distributed online rating task to obtain human estimates of our subjects' social signaling motives. To do so, we randomly created 34 "packets" of 16 user pages – each containing a balanced number of small, medium, quite large and large user pages. We then randomly assigned these packets to 510 independent raters recruited from the online crowdsourcing platform *Prolific.co* in order to obtain 15 independent ratings per subject ($\frac{510\times16}{544}$). We instructed our participants to read the content of each page and rate the extent to which each editor appeared motivated to showcase their skills and achievements to the extended community of contributors.

[11] See https://en.wikipedia.org/wiki/Wikipedia:User page design guide/Introduction

Comprehending the content of a Wikipedia user page requires some level of insider knowledge. In order to guide our online raters through this information, we developed a set of instructions aimed at introducing them to Wikipedia's main social signaling practices.[12] Before deploying our design, we ran our initial version by a sample of students familiar with Wikipedia editing in order to refine our instructions. (See Figure 4, which features the introductory screen of the instructions set.) After reading those instructions, the Prolific raters had to read the 16 user pages assigned to them in turn, and answer the following question: "On a scale from 0 to 10, how much did this editor write their personal page in a way that seeks to explicitly advertise their skills and achievements on Wikipedia?" We averaged the resulting 15 independent scores at the editor level to derive our baseline measure of subjects' relative taste for social "signaling" or image. (Subjects who did not create their user page or decided to leave it blank received a social signaling score of zero.)

[FIGURE 4 ABOUT HERE]

We launched our distributed user page rating study in October 2023. It took 30 minutes to complete on average, paid a fixed fee of £4.50 (i.e., about $5.75), and only recruited from countries where English is the dominant language. Our design screened participants who failed to load all the user pages in their packet, or who gave the same rating to all.[13]

2.3.2 Barnstar signaling in the field

In order to complement the social signaling scores we obtained from our distributed online "biography" rating task, we relied on the data available from Wikipedia to elicit social image prefer-

[12] In Appendix C, we provide a detailed account of these rating procedures.

[13] In practice, the social signaling scores we obtain from this distributed online rating task might be endogenous to subjects' editing activity on Wikipedia. In other words, editors who contribute significantly more content might receive higher social signaling scores, not because of their taste for social image, but simply because they have more achievements to advertise on their personal user pages. In our regressions, we account for this possibility by systematically controlling for (i) the sheer *size* of subjects' user pages, and (ii) the overall number of achievement awards (i.e., Barnstars) they received from the community (see Section 4.1). Further, we address this concern directly in the next Section by looking at whether, conditional on receiving Barnstars, editors take manual steps to make these awards more visible to others, thus explicitly pointing at social signaling motives.

ences based on observed field behavior. Namely, we exploited the fact that Barstars are typically posted on the *talk page* of the recipient which – unlike their personal user or "biography" page – is a popular and convenient place for editors to communicate with one another (i.e., send messages, request help, ask questions and coordinate work). Recently awarded Barnstars thus appear within the flow of conversations between the target contributor and the rest of the community. Eventually, the talk page thread where the Barnstar got posted will be archived and/or become too long for anyone to easily notice that the editor received a community award.

However, similar to, e.g., Congressional Gold Medal recipients who decide to prominently exhibit or wear their medals as opposed to leaving them in their office drawer, some Wikipedia contributors take action to safeguard and advertise their Barnstars by manually moving them to a dedicated "gallery of awards" section of their personal user page. While limited to superstar editors, this practice therefore allows us to identify those who reveal a strong relative taste for social image within the community of contributors. Hence, among subjects who received Barnstars in our sample (the "superstars"), we coded as "social signalers" those who decided to display at least one of their awards on their personal user page.[14] We use the resulting binary variable as an alternative indicator of social image concerns based on observed field behavior.

2.4 Dependent variables: field cooperation

One important contribution of our paper is to empirically account for the fact that, in the field, cooperation is typically a multi-dimensional construct. As a result, public good contributors may, depending on their social motives, arbitrate differently between potentially conflicting dimensions of cooperation at the group level. We therefore study the relationship between social motives and field cooperation by distinguishing three non-overlapping dimensions of cooperation on Wikipedia: (i) the quantity of contributions made, (ii) the level of cooperativeness exhibited towards others while editing, and (iii) the quality of these contributions, which we measure from the time of our experiment on December 8th, 2011 and until November 1st, 2020 (i.e., over 9 consecutive years).[15]

[14] Another empirical strategy could be to take the *proportion* of their community awards that superstar editors choose to display on their personal page as alternative indicator of social image concerns. We show in Appendix E.5 that this variable construction choice does not affect our results.

[15] We discuss possible alternative measurement strategies for our dependent variables in Appendix D.

2.4.1 Quantity of contributions

Measuring the quantity of contributions made to the public good represents the most standard way to operationalize cooperation in the literature, including that on Wikipedia. We follow this literature and define the quantity of contributions made to Wikipedia by our 730 subjects in two complementary ways:

1. The total number of Wikipedia contributions (or "edits") made to the project, defined as the action of (i) going to a Wikipedia page, (ii) hitting the "edit" tab, (iii) implementing a modification, and (iv) saving the modification.
2. The total number of bytes added to the project.

The total number of edits made to Wikipedia by a given contributor is the statistic most frequently tracked by community members to get a sense of their involvement. This statistic is readily available for all registered users, and is sometimes publicized by contributors themselves through dedicated userboxes posted on their user page. Wikipedia even maintains various lists of editors ranked by their edit count (recent or overall).[16] On the other hand, the edit count does not necessarily reflect the amount of original content contributed to the encyclopedia, which is more accurately captured through the total number of bytes added. Both variables should be distributed as power laws in our population of subjects, as skewed contributions is a well-known characteristic of participation in Wikipedia and other digital spaces such as open source software or online message boards (e.g., Stack Overflow, Reddit).

2.4.2 Interpersonal cooperation

Irrespective of the quantity of contributions that individuals choose to make to the public good, their level of cooperativeness with other editors constitutes another important dimension of field cooperation. To be sure, uncivil behavior *while contributing* can be detrimental to overall public goods provision, as such behavior imposes negative externalities on other contributors by increasing the cost of cooperation, and can potentially drive well-intentioned contributors away from the community.

[16] See, e.g., https://en.wikipedia.org/wiki/Wikipedia:List of Wikipedians by number of edits

In order to proxy for the quality of interpersonal cooperation at the subject level, we consider how likely editors are to delete (i.e., "revert") the contributions of others without providing an explanation. Wikipedia editors are strongly encouraged to provide a brief summary for every edit they make, so that other editors could get a quick sense of its purpose. As the "Edit Summary" Wikipedia help page reads: "It is considered good practice to provide a summary for every edit, especially when reverting (undoing) the actions of other editors or deleting existing text."[17] Thus, Wikipedia contributors typically consider non justified reverts as highly uncooperative and harmful to the project.

In an influential paper, Halfaker et al. (2011) collected a sample of 400,000 reverts to understand their effect on subsequent editing behavior. They showed that reverts were especially demotivating for novice contributors, which could be driving Wikipedia's difficulties at retaining new editors. However, novice contributors who survive the reverting process increase the quality of their subsequent edits. In a follow-on paper Geiger et al. (2012), ran field experiments to identify the effect of directed messages aimed at justifying the revert to the reverted editor. They showed that personal messages where the reverting contributors identify themselves and take direct responsibility for rejecting the contribution are most effective at increasing subsequent retention rates. In a related field experiment, Zhu et al. (2013) showed that receiving constructive human feedback of any kind – positive, negative, directive or social – significantly increase editor retention.

Informed by recent behavioral work on gender differences in self-confidence and competitiveness (e.g., Niederle and Vesterlund (2007); Buser et al. (2014)), instances of low interpersonal cooperation while contributing – typically, reverting one's work without providing any explanation – have been argued to be an important driver of the gender gap in the population of editors (Collier and Bear, 2012; Reagle, 2013). We build upon this line of work and proxy for subjects' level of interpersonal cooperativeness while editing by computing the proportion of their reverts on other editors which do not feature an explanation. Out of our 730 subjects, 84 did not edit Wikipedia over the time period we cover, and an additional 108 did not engage in reverting behavior. We therefore recover this indicator for all remaining 538 subjects.

[17] See https://en.wikipedia.org/wiki/Help:Edit summary

2.4.3 Quality of Contributions

In order to distinguish the quantity of contributions made by individual editors from their quality, we build upon a large literature that has studied edit quality on Wikipedia. In a series of early contributions, Adler and De Alfaro (2007), Adler et al. (2008), and Adler et al. (2008) noted the prominence of edit count measures, both in Wikipedia research (as a convenient proxy for cooperation), and as a criterion to recognize merit, award Barnstars, and derive social status within the community of editors. They argued that such measures are "vulnerable to manipulation, and the total text criterion fails to reward people who polish or re-arrange the content" (Adler et al., 2008). Instead, those authors proposed a simple measure of the *quality* of editors' work: how well it survives the peer review of the users who subsequently edit the page. They showed that "content persistence" provided a meaningful indicator of users' editing quality Adler et al. (2008), which may be used to design quality driven reputation systems Adler and De Alfaro (2007), or predict the credibility of a piece of text based on the average persistence of the contributions made by the users who edited it (Adler et al., 2008).

Building upon this work, Halfaker et al. (2009) and Panciera et al. (2009) further explored the concept of content persistence. They measured the average "word persistence" of editors, i.e., the mean number of subsequent article revisions successfully passed by the set of words contributed by an editor.[18] They notably showed that articles edited by "higher quality" contributors – i.e., individuals who, as judged by their peers, wrote higher quality content on average – are more likely to obtain a quality label from the community. Relatedly, Halfaker et al. (2011) used this measure to study the effect of reverts on the quality of editors' subsequent work. Last, Biancani (2014) found that the average word persistence of a given edit was significantly correlated with human quality ratings obtained from *Amazon Mechanical Turk*.

Our design follows this line of work. Each revision made to a Wikipedia article can be retrieved from the history logs, so that the difference between two consecutive versions can be computed. From there, we discretize the text contributed by each subject through each edit into separate words (or "tokens"), and use the concept of word persistence to measure the average

[18] In practice, this indicator represents a refinement and generalization of editors' *revert rate*. See how the concept and method are presented on their dedicated Wikimedia page: https://meta.wikimedia.org/wiki/Research:Content persistence, and https://meta.wikimedia.org/wiki/Research:Measuring edit productivity, respectively.

quality of the contributions made to Wikipedia by our subjects.[19] In practice, our implementation counts the log number of article revisions subsequently performed by *other* editors that each word contributed by the subject survives on average. For instance, an average log persistence of 1 would mean that the average word contributed by the subject survives exp(1) = 2.7 subsequent revisions made by other editors. The interpretation is straightforward: the higher the log persistence, the better the average quality of the contributions made by the subject as judged by the community of editors.

In practice, the size of the corpus of contributions assembled over the 9 consecutive years covered by our study prevents us from defining the concept of log persistence over all subsequent versions of the many articles revised by our subjects. To reduce our computations to a manageable level, we follow the literature and set our program to look at a maximum number of 50 subsequent revisions performed by editors other than the subject on each target article contribution. As a result, for each word w contributed by subject i, $Max[(log\ persistence)_w] = ln(50) = 3.912$ (by construction), and we define the average log persistence of the contributions made by each subject as:

$$Avg(log\ persistence)_i = \frac{\sum_{word=1}^{nb\ words} log(nb\ of\ revisions\ persisted)}{nb\ words} \tag{1}$$

Out of our 730 subjects, 84 subjects made no edits to Wikipedia over our time period. In addition, we could not compute our word persistence indicator for 41 additional subjects with a very low edit count, e.g., because they contributed to pages which got subsequently deleted. We successfully computed our mean word persistence measure of quality for all remaining 605 subjects.

[19]Specifically, we downloaded the full Wikipedia data dumps in the form of compressed XML files to extract the Wikipedia pages edited by our subjects starting from the time of the experiment, on December 8th, 2011, and until November 1st, 2020. As the entire dump was too large to be processed serially (i.e., on a single core), we chunked these files into smaller pieces and processed them simultaneously using multiple CPU cores on the Euler cluster of ETH Zurich. Given the significant editing scope of our subjects, this procedure preserved a large fraction of Wikipedia articles, representing millions of edits. We computed the diffs between the subsequent revisions of our target set of Wikipedia pages using the 'dump2diffs' utility of the 'mwdiffs' tool (see https://pythonhosted.org/mwdiffs/utilities.html). We then processed these diffs to compute the persistence of each added token using the 'persistence2stats' utility of the 'mwpersistence' tool (see https://pythonhosted.org/mwpersistence/utilities.html).

3 Descriptive Data Analysis

3.1 Relationship to the 2011 Wikipedia editor survey

We begin this Section by comparing the demographic characteristics of our subjects pool with that of the 5,281 respondents of the 2011 Wikipedia editor survey. Designed by the Wikimedia Foundation at about the same time at which we ran our experiment, this survey was precisely implemented so as to get a more precise picture of the profiles of Wikipedia editors. Similar to the present study, it was advertized through a Wikipedia banner, and ran for 7 days over the whole population of registered Wikipedia editors.[20]

In Table 1, we compare the demographic information commonly available in both studies (with p-values from t-tests). It appears that the demographic characteristics in both samples are quite similar. This suggests that despite our smaller sample size, we were able to recruit a pool of subjects that is representative of the typical demographic profiles found on Wikipedia, with the average Wikipedia contributor being a 32-33 years old male holding a Bachelors degree. While suggestive, we unfortunately cannot extend this comparison to the other demographic control variables we collected (i.e., salary range and risk aversion level), as the Wikipedia editor surveys purposefully ask relatively few demographic questions in order to maximize the overall response rate for their own target questions of interest (relating to, e.g., user experience and software design issues).

[TABLE 1 ABOUT HERE]

3.2 Sample descriptive statistics

In Table 2, we describe our variables in greater detail. With respect to our dependent variables (top panel), a comparison of the median values of our quantity of contribution measures with the 10th and 90th percentiles of the distributions (reported in brackets) confirms that those variables

[20] See https://meta.wikimedia.org/wiki/Research:Wikipedia Editors Survey 2011 November

are distributed as power laws. Further, Wilcoxon-Mann-Whitney tests of equality in central tendency strongly reject the equality of distributions between casual contributors and superstars for both variables (p <0.001): superstar editors contribute significantly more content than casual ones (e.g, the median number of contributions is 167 over our entire time period for casual contributors, as opposed to 3,222 for superstars).

In terms of interpersonal cooperation, we can see that about a fourth of the reverts performed by our subjects remain unexplained on average. Superstars appear relatively more cooperative than casual contributors in this respect (i.e., 25 vs. 29%, respectively), but this difference is barely statistically significant (t-test, p = 0.094).

Finally, our descriptive statistics with respect to the quality of editors' contributions draw a rather surprising picture. Across all subjects, the average word contributed to Wikipedia survives about exp(2.09) = 8 subsequent page revisions by other editors. This number, however, is significantly lower for superstars than for casual contributors (7 vs. 9.6 revisions respectively, representing a 37% increase in persistence; t-test, p <0.001). All in all, it looks as if superstar contributors contribute significantly more content than regular editors, but do so at the expense of quality.

We further describe our dependent variables in Figure 5, where we report their evolution among both casual and superstar contributors over the 9 consecutive years covered by our study. We can see that superstar contributors maintain significantly higher levels of participation than casual ones over the entire time period. Their participation does tend to level off over time, however, both in terms of number of edits and the amount of content contributed. Casual contributors behave similarly, but sustain more modest, though significant contribution levels over time. Similar to Table 2, we further see that superstar contributors maintain a lower proportion of unjustified reverts over the entire time period, but this difference does not reach statistical significance. Finally, Figure 5 also confirms that superstar editors contribute significantly *lower* quality text than casual ones on average, while both groups see a slight, progressive and parallel decrease in the quality of their contributions as they gain years of seniority.

[FIGURE 5 ABOUT HERE]

Turning our attention back to the middle panel of Table 2 – where we describe our social preferences variables – we first see that the distribution of social types in our subjects pool (based on the conditional public goods game data) delivers results that are qualitatively in line with the laboratory literature (see, e.g., Chaudhuri (2011) or Fallucchi et al. (2019)). In particular, the overwhelming majority of our subjects behave either as full or weak reciprocators (38 and 47%, respectively). The proportion of free-riders (about 7% in our data) does appear lower than the proportion of 20-30% usually obtained with more standard subject pools, however. Similarly, more subjects behave as pure altruists in our data (about 9%). Comparing the frequencies of subjects' social types between the groups of casual and superstar contributors, we can see that the latter contains significantly more altruists (11 vs. 7%, respectively; t-test, $p = 0.034$). The proportions of free-riders and weak reciprocators also appear lower within the group of superstars, but these differences are not statistically significant.

In terms of our social recognition and signaling data, we can see that 47.5% of our subjects have been awarded at least one Barnstar by the extended community of contributors. By definition, all of those subjects are classified as "superstars". Within this group, the mean number of Barnstars received is about 6. As we could expect, our measure of the size of subjects' user pages indicates that superstars maintain significantly larger personal pages than casual ones. Our distributed social signaling rating task further reveals that, as judged by the *content* of their user pages, superstar contributors appear relatively more motivated by social image concerns. Next, we look at the proportion of superstars in our data who chose to manually display at least one of their Barnstars on their personal user page (the "social signalers"). We can see that there is significant heterogeneity in social image concerns even *within* the group of superstar editors, as this proportion is of exactly 50% in our data.

Last, the bottom panel of Table 2 complements Table 1 by providing some more detailed descriptive statistics on our demographic control variables. The average editor in our data is about 33 years old, but superstar contributors are relatively older than casual ones (35 vs. 31 years old, respectively; t-test, $p < 0.001$). Apart from this slight difference, the other variables do not appear to differ significantly between both groups. Only 10% of editors in our data self-identify as female, and the typical contributor has completed between 2 and 4 years of higher education, earns between 1,000 and 3,000 USD per month at the time of the experiment, and exhibits a relatively average level of risk aversion (5.74 out of a 10-points scale).

[TABLE 2 ABOUT HERE]

3.3 Pairwise correlation between the dependent variables

While both Table 2 and Figure 5 clearly show that superstar editors contribute a lot more content than casual ones, our raw data also suggests that superstars tend to contribute lower quality content on average. Table 3 confirms this finding by reporting the pairwise correlations between our dependent variables. Since the distribution of our quantity of contribution measures is skewed, we log-transform those variables as $Ln(1 + \textit{nb contrib.})$ and $Ln(1 + \textit{bytes added})$, which will also ease the interpretation of our regression coefficients as semi-elasticities (see Section 4).

This correlation matrix globally confirms the idea that, at the contributor level, the quantity and the quality of editors' contributions do not necessarily move together. Indeed, our log persistence measure is strongly negatively correlated with both the log-number of contributions made and the log-number of bytes added to the project ($corr =$ -0.203 and -0.278, respectively, with $p <$0.001 in both cases). By contrast, our proxy for the quality of interpersonal cooperation, $Prop(\textit{rv w/o expl.})$, is positively and significantly correlated with both the quantity and the quality of contributions made. Those pairwise correlation results thus stress the importance of distinguishing between the various dimensions of cooperation in a given field context, e.g., to be able to understand how people might arbitrate between possibly conflicting goals at the group level as a function of their social motives.

[TABLE 3 ABOUT HERE]

3.4 Are Wikipedia contributors more prosocial?

One important benefit of lab-in-the-field designs is that they generate their data through experimental procedures that allow researchers to make comparisons across populations and contexts

(Gneezy and Imas, 2017). We exploit this feature of our public goods game data to shed a different light on our sample of subjects, namely, by comparing them to a more "standard" subjects pool. By "standard", we do not mean "representative", but simply the pool of subjects that constitutes the behavioral baseline for the overwhelming majority of experimental research: University students from a "western, educated, industrialized, rich and democratic" country (Henrich et al., 2010). We do this for two reasons. First, it is quite unclear what the most relevant field comparison group ought to be in our case. Second, student subjects provide a convenient baseline for field behavioral research, as their decisions in standard game-theoretic paradigms have been extensively studied and documented (also in comparison to other "non standard" field subjects pools).

In order to perform this comparison, we use the data from Hergueux and Jacquemet (2015), who recruited 154 student subjects to participate in an online public goods game using the exact same design as the one we rely on in this paper. Their home-based online experiment closely mirrors our own field setting, and their design launched at about the same time, in November 2011. The students in their sample are, on average, 27 years old (std=10.88), 60% female, completed a 2 years college degree, have a monthly revenue approximating $1,000, and feature an average level of risk aversion (5.42 out of a 10-points scale). We present the result of our comparison in Table 4, where all the regressions include the above list of individual controls (also reported in Table 2).

In column (1), we start by contrasting subjects' overall contributions to the public good by running a linear regression of the average proportion of their endowment which they conditionally contributed over their 11 conditional contribution decisions. The excluded group in this regression – that of casual Wikipedia editors – contributed about 50% of its endowment, on average, across all conditional contribution decisions. By contrast, student subjects appear substantially less prosocial: they conditionally contributed about 12% less of their endowment across all decisions (i.e., a 24% decrease in conditional contributions, p<0.001). When compared to casual editors, Wikipedia's superstars conditionally contribute an additional 4% of their endowment to the public good (i.e., a 8% increase in conditional contributions, p=0.05).

In columns (2)-(5), we seek to refine those results by presenting probit estimates of the probability that subjects fall within each social type, respectively, free-riders, weak reciprocators,

reciprocators and altruists. We find that student subjects are 9% less likely to behave as reciprocators than casual Wikipedia editors (out of a mean value of 37%), and 4.6% less likely to behave as altruists (out of a mean value of 6.5%). The mirror image of these results is that student subjects are more likely to be classified as free-riders or weak reciprocators (but those coefficients are not statistically significant). Our estimates further suggest that superstar editors are more likely to behave as reciprocators or altruists than casual contributors (+2 and +2.5%, respectively) – and, consequently, less likely to be behave as free-riders or weak reciprocators – but these more modest differences fail to reach statistical significance.

Taken together, the results of this comparison of the experimental behavior of our subjects against that of a more traditional subjects pool show that our population of casual Wikipedia editors is significantly more prosocial than standard laboratory samples. Further, our results suggest that Wikipedia's superstars are, on average, more prosocial than casual editors themselves. With those descriptive results in mind, Section 4, turns to the question of the empirical *relationship* between social preferences and field cooperation.

[TABLE 4 ABOUT HERE]

4 Social Motives and Cooperation on Wikipedia

In this Section, we rely on our full sample of Wikipedia subjects to estimate the average relationship between social preferences and field cooperation. We then interact our social motive variables of interest with subjects' superstar status in order to test for differences in those average relationships (Section 4.1). In a second step, we focus our analysis on the Wikipedia superstars, and use our field measure of social signaling to extend and confirm the results of our analysis (Section 4.2).

4.1 The relationship between social motives and cooperation

We estimate the relationship between social preferences and field cooperation by running the following regression at the subject level:

$$
\begin{aligned}
Cooperation_i = \beta_0 + \beta_1 Reciprocator + \beta_2 Altruist + \beta_3 Social\_signaler \\
+\gamma Controls + \delta_1 Nb\_Barnstars + \delta_2 Size\_Userpage + u_i,
\end{aligned}
\tag{2}
$$

where $Cooperation_i$ is either:

$$
\begin{cases}
(1) Ln(1+ \textit{ nb contrib.}) : \text{our first measure of quantity of contributions,} \\
(2) Ln(1+ \textit{ bytes added}) : \text{our second measure of quantity of contributions,} \\
(3) Prop(\textit{rv w/o expl.}) : \text{our proxy for interpersonal cooperation,} \\
(4) Avg(\textit{log persist.}) : \text{our measure of quality of contributions,}
\end{cases}
$$

Controls stands for the set of individual control variables reported in Table 2, *Nb_Barnstars* is the overall number of Barnstars received by the subject, and *Size_Userpage* is the size of their "biography" page (in bytes). Because the proportion of free-riders is relatively low in our experiment (i.e., 7%), we group this social type with that of weak reciprocators to form the baseline category in the above specification.[21]

In this baseline specification, *Social_signaler* is proxied by the mean social signaling score received by subjects in the online user page rating task. Because this regression controls for *Nb_Barnstars* and *Size_Userpage*, the coefficient β_3 is identified *within* strata of contributors who received the same number of community awards and have a similar-sized user page, but differ in terms of their social signaling preferences. In a second version of this baseline specification, we interact our three social preference variables of interest with subjects' superstar status in order

[21] In Appendix E.3, we reproduce our analysis by trying to single-out free-riders as the baseline category in our estimations. We show that our limited sample size makes it difficult for our model to estimate an additional social type coefficient without imposing a prohibitive cost in terms of the statistical power required for our tests. However, this strategy does generate results that are consistent with those reported in the main text when we implement a less stringent social type classification rule, which allows subjects for some modest amount of decision error in our experiment (thus increasing the number of subjects falling in the baseline free-riding category).

to allow for differences in the relationship between social motives and field cooperation in both subgroups.

We present our results in Table 5.[22] We first focus on columns (1)-(4), where we estimate our baseline specification on the full sample of Wikipedia subjects. On average, full reciprocators in the public goods experiment make 58% more edits than weak reciprocators and free-riders over our entire time period (column (1)).[23] They also contribute about 96% more content (column (2)), and appear significantly more cooperative while editing, as they leave 4.7% less of their reverts unexplained (column (3)) out of a mean value of 26% (i.e., a 18% decrease in proportion). Altruist editors follow a similar pattern – they notably contribute twice as much content as their less prosocial counterparts (column (2)) – although our available sample size appears to limit our ability to precisely estimate some of these coefficients (columns (1) and (3)).

By contrast to both reciprocal and altruistic preferences, the relationship between social image motives, as revealed by our online user page rating task, and the quantity of contributions made appears an order of magnitude stronger. Specifically, the coefficient we estimate on social signalers is about 12 times (respectively, 14 times) as large as that on reciprocity preferences, with respect to both the number of edits made (column (1)) and the amount of content contributed (column (2)). In this baseline model, however, social image motives do not appear related to our measure of cooperativeness while editing (column (3)).

Turning our attention to our measure of contribution quality (column (4)), we see no relationship with either reciprocal or altruistic preferences. Our results with respect to social image motives, however, resonate with our descriptive findings from Table 3, where we uncovered a surprising *negative* correlation between our measures of contribution quantity and quality at the editor level. Namely, the social signalers in our data, if they contribute significantly more content to Wikipedia, also contribute *lower* quality material on average. In practice, this means that, as

[22]In this table, the number of observations per column differ depending on the set of subjects for whom we could compute each of our indicators of field cooperation (as described in Sections 2.4.1, 2.4.2 and 2.4.3, respectively), and who answered all our control survey questions. In Appendix E.1, we report the result of our estimations without control variables. This allows us to explicitly discuss the scope of *observable* omitted variables bias in our model, and present a few informative statistics with respect to the possible scope of *unobserved* omitted variables bias (Oster, 2019). By contrast, in Appendix E.2, we report the results of our estimations with our full set of control variables, but restrict the sample across regressions so that it remains stable across columns.

[23]The exact effect size is computed as $e^{\hat{\beta}} - 1$.

vetted by their peers, social signalers contribute content that persists about 38% less revisions on average.[24]

Finally, our control variables are also related to field cooperation in insightful ways. Four interesting patterns emerge across columns (1)-(4). First, older editors appear more cooperative by two of our measures: (i) they tend to contribute significantly more content (columns (1)-(2)), and (ii) they are less likely to leave their reverts unexplained (column (3)). Second, not only do women contributors represent a small minority of Wikipedia editors (10% in our data; see also Collier and Bear (2012) and Hill and Shaw (2013)), they also participate significantly less on average, i.e., they make 38% less edits (column (1)), and contribute 24% less content (column (2)). Third, editors' level of education is strongly associated with the quality of their edits (column (4)). Out of an 8-points scale, each additional degree level yields an average increase of 6% in content persistence. This represents a sizeable number: all else equal, an editor moving from the lowest education level in our data (i.e., who did not complete high school), to the highest (i.e., earned a PhD), would thus see the persistence of their contributions increase by 48% on average.

Finally, and perhaps most interestingly, we can see that the total number of Barnstars received by our subjects is significantly associated with both the quantity of contributions made (columns (1)-(2)), and the level of cooperativeness exhibited towards others (column (3)). However, consistent with the previously made argument that, based on the tools currently deployed within the community of contributors, overall edit quality is both difficult and costly to observe for individual editors (Adler et al., 2008; Adler and De Alfaro, 2007; Adler et al., 2008), we find that the number of community awards received by our subjects is, on average, not related to the quality of the content they contributed (column (4)). Further, we find evidence consistent with the idea that a lack of readily available information about edit quality may drive this result, as individual editors *do* seem to care about the quality of their edits. This can be seen from the fact that, across columns (1)-(4), the size of their user page is uniquely related to our measure of

[24] In Appendix E.6, we present direct empirical evidence ruling out an alternative potential explanation for these results. Editors motivated to earn social recognition might simply engage with relatively more controversial topics on average (e.g., because it would give them more visibility). Such editing behavior may mechanically decrease the overall persistence of the content contributed by these subjects, thus potentially inducing omitted variables bias on our social preference estimates of interest in Tables 5 and 6. Building upon a separate stream of the Wikipedia literature which has sought to estimate the level of conflictuality of Wikipedia articles (Yasseri et al., 2012, 2014; Chhabra et al., 2020; Greenstein et al., 2021), we show empirically that this is not the case.

contribution quality.

As a second step to this analysis, we interact our three social preference variables of interest with subjects' superstar status, to test for systematic differences in the above relationships in both subgroups. We report our results in columns (5)-(8) of Table 5. Focusing first on our measures of cooperativeness while editing (column (7)), and edit quality (column (8)), it appears that the limits of our study in terms of statistical power likely prevent us from further refining the average results reported in columns (3)-(4). Specifically, the estimated relationship between reciprocal preferences and our measure of individual cooperativeness while editing (column (7)) remains negative in both subgroups, but does not reach statistical significance anymore. Similarly, the relationship between social image preferences and our measure of edit quality (column (8)) remains negative for both casual and superstar editors, but now loses statistical significance.

By contrast, a striking pattern emerges across columns (5)-(6), which analyze the quantity of contributions made by subjects. First, our estimates from columns (1)-(2) of the average relationship between reciprocity, social image preferences and contribution quantity appear largely driven by the group of casual editors. When focusing on this subgroup, the magnitude of our coefficients on reciprocal motives increase by 74% (from column (1) to column (5)) and 50% (from column (2) to column (6)). Similarly, within the group of casual contributors, the magnitude of our coefficients on social signaling motives increase by 43% (from column (1) to column (5)) and 50% (from column (2) to column (6)).

Second, for both contribution quantity variables in columns (5)-(6), our model reports a strong negative interaction between reciprocity and social image preferences, on the one hand, and superstar status, on the other. This result represents the mirror image of the increased magnitudes discussed above in the context of casual Wikipedia editors. In particular, the size of these interactions are such that the resulting total coefficients within the group of superstar contributors become close to zero for reciprocity preferences, while they remain positive, though significantly smaller, for social image preferences. An empirical analysis focused on superstar editors therefore appears warranted before one can conclude, with more statistical confidence, with respect to the precise nature of the relationship between social preferences and field cooperation within this group.

[TABLE 5 ABOUT HERE]

4.2 The heterogeneity of motives: Wikipedia's superstars

In this Section, we focus our analysis on the Wikipedia superstars by restricting our baseline specification to this subgroup of subjects. Within this group, we collected an additional, field-based measure of social signaling preferences, based on whether editors decided to prominently display (at least one) of their Barnstars on their personal user page. As a result, the variable *Social_signaler* can now take two values in Equation (2), and we test both in turn in order to explore the consistency of our estimates. In both cases, we continue to control for *Nb_Barnstars* and *Size_Userpage* in our regressions, so that the coefficient on *Social_signaler* is identified within groups of contributors that received the same number of Barnstars, and feature similar-sized user pages.

We present our results in Table 6. A clear picture emerges: consistent with our results from Table 5, within the group of superstar editors, reciprocity preferences do not explain inter-individual differences in contribution quantity (columns (1)-(4)). It is worth noting, however, that altruistic types within the group of superstars appear significantly more cooperative towards others when reverting their contributions, as they leave about 9% less of their reverts unjustified (columns (5)-(6)). Out of a mean value of 25%, this represents a 36% decrease in proportion.

By contrast to reciprocity preferences, social image motives remain strongly associated with the quantity of contributions made among superstar editors (columns (1)-(4)). Depending on the social signaling measure we use, we see from columns (1)-(2) that social signalers within this group contribute between 2 and 4.5 times more edits than non social signalers. Similarly, based on columns (3)-(4), they contribute between 1.5 and 7.5 more content on average. Given that superstar contributors typically sustain significantly higher participation levels – the median number of edits among casual editors in our data is 167, as compared to 3,222 for superstars (see Table 2) – these coefficients translate into sizable differences in contribution quantity.

Finally, the estimates from columns (7)-(8) confirm that superstar editors who reveal a relatively stronger taste for social image, if they produce more edits, contribute lower quality content on average. Specifically, within the group of superstars, social signalers contribute content that persists between 26 and 34% less subsequent peer revisions on average, depending on the social signaling variable we use.

Our above analysis of the relationship between social preferences and field cooperation among superstar contributors thus reveals a striking contrast. While reciprocity and social image preferences both appear strongly related to the quantity of field contributions made among casual contributors, only social image concerns remain associated with systematic differences in contribution quantity among the Wikipedia superstars. Further, consistent with the average relationships we uncovered in Tables 3 and 5, we find that superstar editors who reveal a relatively stronger taste for social image behave as if they were substituting quality for quantity in their contribution decisions. As we already pointed out, measures of overall edit quality are not easily available at the editor level, and require an in-depth, manual exploration of their contribution records. The likely result of this measurement issue on social rewarding practices – i.e., the fact that Barnstars appear largely awarded as a function of contribution quantity and individual cooperativeness, but not content quality – might thus create a perverse incentive for contributors motivated to earn social recognition, leading them to actively favor quantity, at the expense of quality.

[TABLE 6 ABOUT HERE]

5 Social Motives and Cooperation: Dynamic Analysis

The results presented so far in this paper were based on the average of our dependent variables over the entire time period covered by our study – from the time of our experiment on December 8th, 2011 and until November 1st, 2020 (i.e., 9 consecutive years). In this section, we discretize our dependent variables on a yearly basis, i.e., from +1 to +9 years after the elicitation of our independent variables. We then proceed to study whether and how the average relationships we uncovered between social preferences and field cooperation evolve over time within the groups of casual and superstar editors, respectively.

In order to do that, we rerun the baseline specification reported in Equation (2) separately for year +1, +2, ..., +9 after the experiment, and distinguishing casual from superstar editors. We then extract from each model our main coefficients of interest – that on reciprocity and social image motives – which we plot as a function of time.[25]

We present our results in Figure 6, where we distinguish casual contributors (left) from superstars (right). We can see from the first and second rows that the association between reciprocity, social image preferences and the quantity of contributions made by casual contributors is fairly stable over the entire time period, even though, for the former variable, the 95% confidence interval around our estimates contains zero in some of the last periods (i.e., years +8 and +9 after the elicitation of our independent variables). Consistent with the extant literature (e.g., Carlsson et al. (2014)) this suggests that social preferences and their relationship to field behavior are stable over fairly long periods of time.

Our dynamic analysis also confirms that this positive association between reciprocal preferences and contribution quantity does not hold anymore within the group of superstar editors. Rather, in this group, social image concerns display a strong relationship with the quantity of edit and content contributions. Further, the magnitude of this relationship increases significantly as a function of time, as measured by both our online user page rating task, and our field indicator of Barnstar signaling. In other words, superstar editors who reveal strong social signaling preferences make more edits than others on average, and this difference grows larger over time.

Turning our attention to the third row of Figure 6, we can see that, consistent with our results from Tables 5 and 6, the association between reciprocal preferences, social image motives and subjects' level of cooperativeness while editing is unstable, and not well estimated in our data.

By contrast, a dynamic analysis of the evolution of the average quality of the content con-

[25]Since our coefficients on altruistic preferences are much less precisely estimated, we report these coefficients separately in Appendix F. Our results confirm that the relationship between altruism and field cooperation is unstable and not precisely estimated in our data. One exception is worth noting: consistent with our results from Table 6, we do find suggestive evidence that, over our entire time period, altruistic preferences are positively related to superstar editors' level of interpersonal cooperativeness while editing.

tributed by our subjects as a function of reciprocity and social image preferences, which we present in the fourth row, provides interesting insights. Both pictures confirm that individual contribution quality is unrelated to reciprocal preferences. However, editors who reveal a relatively stronger taste for social image – both casual and superstar – consistently contribute lower quality content over our entire time period. This relationship appears particularly stronger within the group of superstar contributors, where it remains stable and statistically significant over the 9 consecutive years covered by our study, than within that of casual editors, where the 95% confidence interval around our estimates contains zero starting from +7 years after the elicitation of our independent variables.

Taken together, these results reinforce our above interpretation that editors who reveal a stronger taste for social image within the community of contributors – both casual and superstar – are likely subject to a perverse incentive effect. As editors mainly earn public recognition by being cooperative towards others and accumulating large edit counts, the social rewarding structure of Wikipedia appears to incentivize social image driven contributors to actively favor quantity, at the expense of quality.

[FIGURE 6 ABOUT HERE]

6 Conclusion

In this paper, we report on a lab-in-the-field study of the relationship between social preferences and field cooperation in Wikipedia – one of Internet's most valuable public good, and largest online volunteering community. On December 8th, 2011, we conducted an online public goods game with a diverse sample of 730 Wikipedia contributors, from which we recovered their reciprocal and altruistic preferences. In addition, we conducted a crowdsourced online study of their Wikipedia personal user (or "biography") pages, to obtain a subject level measure of social image concerns, which we cross-validate within the sample of superstar contributors with a related field measure (based on community award signaling behavior).

Our lab-in-the-field study contributes to the vast literature on public goods provision, and our general understanding of Wikipedia as an online volunteering or "peer production" community, through three distinct features of our design. First, we account for the fact that many volunteering communities organize themselves around two layers of participation. "Core" contributors exhibit quite extreme levels of dedication and involvement with the common project, and play a key role in the functioning of the community. Articulated around them, a large majority of "peripheral" contributors make more discrete contributions, but those can nonetheless add up.

We take this often-observed feature of participation into account in our analysis by distinguishing core community members – the "superstars" – in our sample of subjects. To do so, we let the community point us at their most notable peers, and relied on its main social recognition tool — community awards called "Barnstars" -– to identify the Wikipedia superstars. We show that the relationship between social preferences and field cooperation on Wikipedia is different in both subgroup. Schematically, reciprocal and social image preferences are strongly related to contribution quantity among casual contributors. However, while superstar contributors do appear more prosocial on average, only social image motives continue to predict differences in contribution levels within this group.

Second, our design is built around the observation that field cooperation is typically a multi-dimensional construct, so that contributors usually pursue several (and possibly conflicting) cooperative goals at the group level. For instance, in a given field context, some types of cooperative behavior are usually more visible and/or socially rewarded than others. Contributors motivated by social image concerns might thus arbitrate differently between these concurrent dimensions of cooperation. Similar to monetary incentives, social motives may therefore induce substitution (or "perverse incentive") effects between potentially conflicting dimensions of field cooperation.

We take the multi-dimensional aspect of field cooperation into account in our study by distinguishing between three separate but complementary dimensions of cooperation on Wikipedia: (i) the quantity of contributions made, (ii) the quality of these contributions, and (iii) the level of cooperativeness exhibited towards others while editing. Our data shows that altruistic superstars are significantly more cooperative towards others while editing. Importantly, we find that

Wikipedia editors who reveal a strong relative taste for social image – both casual and superstar – if they make more edits, contribute significantly *lower* quality content on average (as judged by their fellow editors). We present empirical evidence indicative of the fact that social image driven contributors actively favor quantity in their contribution decisions – a more socially visible activity in the community – at the expense of quality.

Third, we conduct our lab-in-the-field study over a 9 years horizon – i.e., from December 2011 and until November 2020 – which allows us to assess the stability of the relationships we uncover between social preferences and field cooperation over a relatively long period of time. Overall, we find these relationships to be remarkably stable. Within the group of casual editors, social image preferences remain strongly related to contribution quantity over the entire time period. The same pattern applies to the (positive) relationship between reciprocal preferences and contribution quantity, and the (negative) relationship between social image preferences and contribution quality (but at the very end of our time period, where theses coefficients lose statistical significance). Within the group of superstar editors, social image concerns are related to contribution quantity at a positive and *increasing* rate over time, while they remain consistently associated with lower levels of contribution quality over the entire time period.

The limitations of our design also suggest fruitful avenues for future research. For instance, compared to casual contributors, superstar Wikipedia editors appear relatively more prosocial in our data. Among superstar editors, field cooperation is further related to social image concerns and altruistic preferences. While suggestive, those findings do not explain why these "superstar types" elected to engage in extreme contributing behavior in *this particular volunteering context*, as opposed to any other. In order to account for such self-selection, economic theory may need to endogenize some of the mechanisms of *identity formation* identified by the extant literature in sociology and social psychology, since those are likely antecedents of the heterogeneity in social image concerns that we document within Wikipedia (Turner et al., 2010; Hogg and Terry, 2014). Recent economic theory research has started to incorporate elements of social identity (see Akerlof and Kranton (2000), Akerlof and Kranton (2010) or Carvalho (2016)). Similarly, experimental research has already demonstrated that social identification is a powerful mechanism to foster prosocial (Chen and Li, 2009), but also anti-social behavior Cohn et al. (2014, 2015). However, the process by which such identification occurs in field settings remains vastly understudied in economics (Kranton, 2016; Charness and Chen, 2020).

References

Aaltonen, A. and S. Seiler (2016). Cumulative growth in user-generated content production: evidence from wikipedia. *Management Science 62*(7), 2054–2069.

Adler, B. T., K. Chatterjee, L. De Alfaro, M. Faella, I. Pye, and V. Raman (2008). Assigning trust to wikipedia content. In *Proceedings of the 4th International Symposium on Wikis*, pp. 1–12.

Adler, B. T. and L. De Alfaro (2007). A content-driven reputation system for the wikipedia. In *Proceedings of the 16th international conference on World Wide Web*, pp. 261–270.

Adler, B. T., L. De Alfaro, I. Pye, and V. Raman (2008). Measuring author contributions to the wikipedia. In *Proceedings of the 4th International Symposium on Wikis*, pp. 1–10.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Akerlof, G. A. and R. E. Kranton (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy 97*(6), 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–55.

Barr, A. and P. Serneels (2009). Reciprocity in the workplace. *Experimental Economics 12*(1), 99–112.

Bartling, B., E. Fehr, and K. M. Schmidt (2012). Screening, competition, and job design: Economic origins of good jobs. *American Economic Review 102*(2), 834–64.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–1093.

Bénabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies 70*(3), 489–520.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Benkler, Y. (2002). Coase's penguin, or, linux and "the nature of the firm". *Yale Law Journal*, 369–446.

Benkler, Y. (2008). *The Wealth of Networks*. Yale University Press.

Benkler, Y., A. Shaw, and B. M. Hill (2015). Peer production: A form of collective intelligence. *Handbook of collective intelligence 175*.

Biancani, S. (2014). Measuring the quality of edits to wikipedia. In *Proceedings of The International Symposium on Open Collaboration*, pp. 1–3.

Bryant, S. L., A. Forte, and A. Bruckman (2005). Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 ACM International Conference on Supporting Group Work*, pp. 1–10.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science 6*(1), 3–5.

Burlando, R. M. and F. Guala (2005). Heterogeneous agents in public goods experiments. *Experimental Economics 8*(1), 35–54.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The quarterly journal of economics 129*(3), 1409–1447.

Carlsson, F., O. Johansson-Stenman, and P. K. Nam (2014). Social preferences are stable over long periods of time. *Journal of Public Economics 117*, 104–114.

Carpenter, J. and C. K. Myers (2010). Why volunteer? evidence on the role of altruism, image, and incentives. *Journal of Public Economics 94*(11-12), 911–920.

Carpenter, J. and E. Seki (2011). Do social preferences increase productivity? field experimental evidence from fishermen in toyama bay. *Economic Inquiry 49*(2), 612–630.

Carvalho, J.-P. (2016). Identity-based organizations. *American Economic Review 106*(5), 410–14.

Charness, G. and Y. Chen (2020). Social identity, group behavior, and teams. *Annual Review of Economics 12*, 691–713.

Charness, G. and C.-L. Yang (2014). Starting small toward voluntary formation of efficient large groups in public goods provision. *Journal of Economic Behavior & Organization 102*, 119–132.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics 14*(1), 47–83.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Chhabra, A., R. Kaur, and S. Iyengar (2020). Dynamics of edit war sequences in wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration*, pp. 1–10.

Chmielewski, M. and S. C. Kucker (2020). An mturk crisis? shifts in data quality and the impact on study results. *Social Psychological and Personality Science 11*(4), 464–473.

Cinyabuguma, M., T. Page, and L. Putterman (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics 89*(8), 1421–1435.

Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature 516*(7529), 86–89.

Cohn, A., M. A. Maréchal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies 82*(4), 1289–1308.

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

Crowston, K., K. Wei, J. Howison, and A. Wiggins (2008). Free/libre open-source software development: What we know and what we do not know. *ACM Computing Surveys (CSUR) 44*(2), 1–35.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Douglas, B. D., P. J. Ewell, and M. Brauer (2023). Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. *Plos one 18*(3), e0279720.

Ellingsen, T. and M. Johannesson (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review 98*(3), 990–1008.

Ellingsen, T. and M. Johannesson (2011). Conspicuous generosity. *Journal of Public Economics 95*(9-10), 1131–1143.

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Fallucchi, F., R. A. Luccasen, and T. L. Turocy (2019). Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association 5*(2), 238–254.

Fehr, E. and G. Charness (2023). Social preferences: fundamental characteristics and economic consequences.

Fehr, E. and A. Leibbrandt (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics 95*(9-10), 1144–1155.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fischbacher, U. and S. Gächter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review 100*(1), 541–56.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Gächter, S. and C. Thöni (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association 3*(2-3), 303–314.

Gallus, J. (2017). Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia. *Management Science 63*(12), 3999–4015.

Geiger, R., A. Halfaker, M. Pinchuk, and S. Walling (2012). Defense mechanism or socialization tactic? improving wikipedia's notifications to rejected contributors. In *Proceedings of the International AAAI Conference on Web and Social Media*, Volume 6, pp. 122–129.

Gneezy, U. and A. Imas (2017). Lab in the field: Measuring preferences in the wild. In *Handbook of Economic Field Experiments*, Volume 1, pp. 439–464. Elsevier.

Gneezy, U., A. Leibbrandt, and J. A. List (2016). Ode to the sea: Workplace organizations and norms of cooperation. *The Economic Journal 126*(595), 1856–1883.

Greenstein, S., G. Gu, and F. Zhu (2021). Ideology and composition among an online crowd: Evidence from wikipedians. *Management Science 67*(5), 3067–3086.

Greenstein, S. and F. Zhu (2018). Do experts or crowd-based models produce more bias? evidence from encyclopædia britannica and wikipedia. *Mis Quarterly*.

Halfaker, A. and R. S. Geiger (2020). Ores: Lowering barriers with participatory machine learning in wikipedia. *Proceedings of the ACM on Human-Computer Interaction 4*(CSCW2), 1–37.

Halfaker, A., A. Kittur, R. Kraut, and J. Riedl (2009). A jury of your peers: quality, experience and ownership in wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pp. 1–10.

Halfaker, A., A. Kittur, and J. Riedl (2011). Don't bite the newbies: how reverts affect the quantity and quality of wikipedia work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pp. 163–172.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). The weirdest people in the world? *Behavioral and brain sciences 33*(2-3), 61–83.

Henry, E. and J. Sonntag (2019). Measuring image concern. *Journal of Economic Behavior & Organization 160*, 19–39.

Hergueux, J., Y. Algan, Y. Benkler, and M. Fuster-Morell (2021). Do i trust this stranger? generalized trust and the governance of online communities. In *Proceedings of the Web Conference 2021*, pp. 539–543.

Hergueux, J., E. Henry, Y. Benkler, and Y. Algan (2023). Social exchange and the reciprocity roller coaster: Evidence from the life and death of virtual teams. *Organization Science 34*(6), 2296–2314.

Hergueux, J. and N. Jacquemet (2015). Social preferences in the online laboratory: a randomized experiment. *Experimental Economics 18*(2), 251–283.

Hergueux, J., N. Jacquemet, S. Luchini, and J. F. Shogren (2022). Leveraging the honor code: Public goods contributions under oath. *Environmental and Resource Economics 81*(3), 591–616.

Hergueux, J. and S. Kessler (2022). Follow the leader: Technical and inspirational leadership in open source software. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.

Hill, B. M. and A. Shaw (2013). The wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PloS one 8*(6), e65782.

Hinnosaar, M., T. Hinnosaar, M. E. Kummer, and O. Slivko (2022). Externalities in knowledge production: evidence from a randomized field experiment. *Experimental Economics 25*(2), 706–733.

Hogg, M. A. and D. J. Terry (2014). *Social identity processes in organizational contexts*. Psychology Press.

Holländer, H. (1990). A social exchange approach to voluntary cooperation. *American Economic Review*, 1157–1167.

Hong, F., L. Karp, and T.-H. Teh (2021). Identity in public goods contribution. *Social Choice and Welfare 57*(3), 617–664.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics 14*, 399–425.

Jacquemet, N., A. G. James, S. Luchini, J. J. Murphy, and J. F. Shogren (2021). Do truth-telling oaths improve honesty in crowd-working? *PloS one 16*(1), e0244958.

Jacquemet, N., R.-V. Joule, S. Luchini, and J. F. Shogren (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management 65*(1), 110–132.

Jacquemet, N., S. Luchini, J. Rosaz, and J. F. Shogren (2019). Truth telling under oath. *Management Science 65*(1), 426–438.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review 95*(5), 1688–1699.

Kittur, A., E. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web 1*(2), 19.

Kosfeld, M. and D. Rustagi (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *American Economic Review 105*(2), 747–83.

Kosfeld, M. and F. A. Von Siemens (2011). Competition, cooperation, and corporate culture. *The RAND Journal of Economics 42*(1), 23–43.

Kranton, R. E. (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review P&P 106*(5), 405–09.

Kriplean, T., I. Beschastnikh, and D. W. McDonald (2008). Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work*, pp. 47–56.

Laniado, D., A. Kaltenbrunner, C. Castillo, and M. F. Morell (2012). Emotions and dialogue in a peer-production community: the case of wikipedia. In *proceedings of the eighth annual international symposium on wikis and open collaboration*, pp. 1–10.

Leibbrandt, A. (2012). Are social preferences related to market performance? *Experimental Economics 15*(4), 589–603.

Lerner, J. and J. Tirole (2002). Some simple economics of open source. *The Journal of Industrial Economics 50*(2), 197–234.

Levine, S. S., O. Schilke, O. Kacperczyk, and L. G. Zucker (2023). Primer for experimental methods in organization theory. *Organization Science*.

Li, A., Z. Yao, D. Yang, C. Kulkarni, R. Farzan, and R. E. Kraut (2020). Successful online socialization: lessons from the wikipedia education program. *Proceedings of the ACM on Human-Computer Interaction 4*(CSCW1), 1–24.

Malmendier, U. and G. Tate (2009). Superstar ceos. *The Quarterly Journal of Economics 124*(4), 1593–1638.

McDonald, D. W., S. Javanmardi, and M. Zachry (2011). Finding patterns in behavioral observations by automatically labeling forms of wikiwork in barnstars. In *Proceedings of the 7th international symposium on wikis and open collaboration*, pp. 15–24.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics 122*(3), 1067–1101.

Olson, M. (2009). *The logic of collective action*, Volume 124. Harvard University Press.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Page, T., L. Putterman, and B. Unel (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal 115*(506), 1032–1053.

Panciera, K., A. Halfaker, and L. Terveen (2009). Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the 2009 ACM International Conference on Supporting Group Work*, pp. 51–60.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making 5*(5), 411–419.

Peer, E., L. Brandimarte, S. Samat, and A. Acquisti (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of experimental social psychology 70*, 153–163.

Peer, E., D. Rothschild, A. Gordon, Z. Evernden, and E. Damer (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 1281–1302.

Reagle, J. (2013). "free as in sexist?" free culture and the gender gap. *First Monday*.

Restivo, M. and A. Van De Rijt (2012). Experimental study of informal rewards in peer production. *PloS one 7*(3), e34358.

Restivo, M. and A. van de Rijt (2014). No praise without effort: experimental evidence on how rewards affect wikipedia's contributor community. *Information, Communication & Society 17*(4), 451–462.

Rustagi, D., S. Engel, and M. Kosfeld (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Science 330*(6006), 961–965.

Safadi, H., S. L. Johnson, and S. Faraj (2021). Who contributes knowledge? core-periphery tension in online innovation communities. *Organization Science 32*(3), 752–775.

Sajnani, H., S. Javanmardi, D. W. McDonald, and C. V. Lopes (2011). Multi-label classification of short text: A study on wikipedia barnstars. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Shaw, A. and B. M. Hill (2014). Laboratories of oligarchy? how the iron law extends to peer production. *Journal of Communication 64*(2), 215–238.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature 43*(2), 392–436.

TeBlunthuis, N., B. M. Hill, and A. Halfaker (2021). Effects of algorithmic flagging on fairness: quasi-experimental evidence from wikipedia. *Proceedings of the ACM on Human-Computer Interaction 5*(CSCW1), 1–27.

Thomas, K. A. and S. Clifford (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior 77*, 184–197.

Turner, J. C., K. J. Reynolds, et al. (2010). The story of social identity. In *Rediscovering social identity: Key readings*. Psychology Press, Taylor & Francis.

van Leeuwen, B., T. Offerman, and A. Schram (2020). Competition for status creates superstars: An experiment on public good provision and network formation. *Journal of the European Economic Association 18*(2), 666–707.

Viégas, F. B., M. Wattenberg, and M. M. McKeon (2007). The hidden order of wikipedia. In *International Conference on Online Communities and Social Computing*, pp. 445–454. Springer.

Wu, F., D. M. Wilkinson, and B. A. Huberman (2009). Feedback loops of attention in peer production. In *2009 International Conference on Computational Science and Engineering*, Volume 4, pp. 409–415. IEEE.

Yasseri, T., A. Spoerri, M. Graham, and J. Kertész (2014). The most controversial topics in wikipedia. *Global Wikipedia: International and cross-cultural issues in online collaboration 25*, 25–48.

Yasseri, T., R. Sumi, A. Rung, A. Kornai, and J. Kertész (2012). Dynamics of conflicts in wikipedia. *PloS one 7*(6), e38869.

Zhang, X. M. and F. Zhu (2011). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review 101*(4), 1601–15.

Zhu, H., A. Zhang, J. He, R. E. Kraut, and A. Kittur (2013). Effects of peer feedback on contribution: a field experiment in wikipedia. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2253–2262.

Figures and Tables

FIGURE (1) The Wikipedia recruitment banner

FIGURE (2) The decision screen of the conditional public goods game

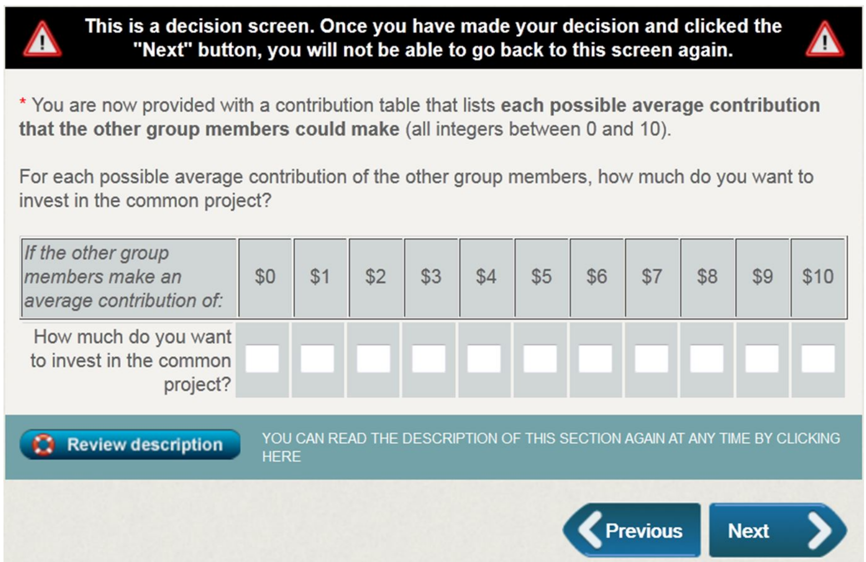


FIGURE (3) A typical Barnstar

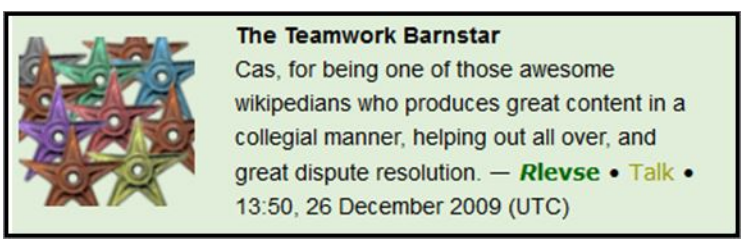


FIGURE (4) The introduction screen of the online instructions for the social image rating task

Detailed instructions (1/5)

Introduction

In this study, we will ask you to **read the personal page of 16 volunteer Wikipedia editors**.

Based on your reading, your task is to **rate the extent to which each editor appears motivated to showcase their skills and achievements** to the extended community of Wikipedia editors.

This is **not** the same thing as rating the length of this editor page. Some editors might post relatively long descriptive (or technical) content on their editor page whose purpose is not to signal the significance of their contributions to the reader. Examples include working drafts of future article contributions, lists of tools and references for later use, personal thoughts on various topics…

In such cases, you do not necessarily need to read in detail all the text posted by each editor to perform your task well. Rather, you can focus on **identifying whether – and to which extent – the editor appears motivated to signal their skills and achievements on Wikipedia**.

To do so, you can read through the text of each page to identify a few **key indicators of social signaling** which may (or may not) be featured on the page. Those indicators generally (but not exclusively) fall within three categories:

1) List of barnstars received (community awards)

2) List of articles edited (together with their quality rating)

3) Userboxes that advertise one's activity and achievements on Wikipedia

Next

FIGURE (5) Dynamics of Wikipedia Contributions: Casual Contributors vs. Superstars

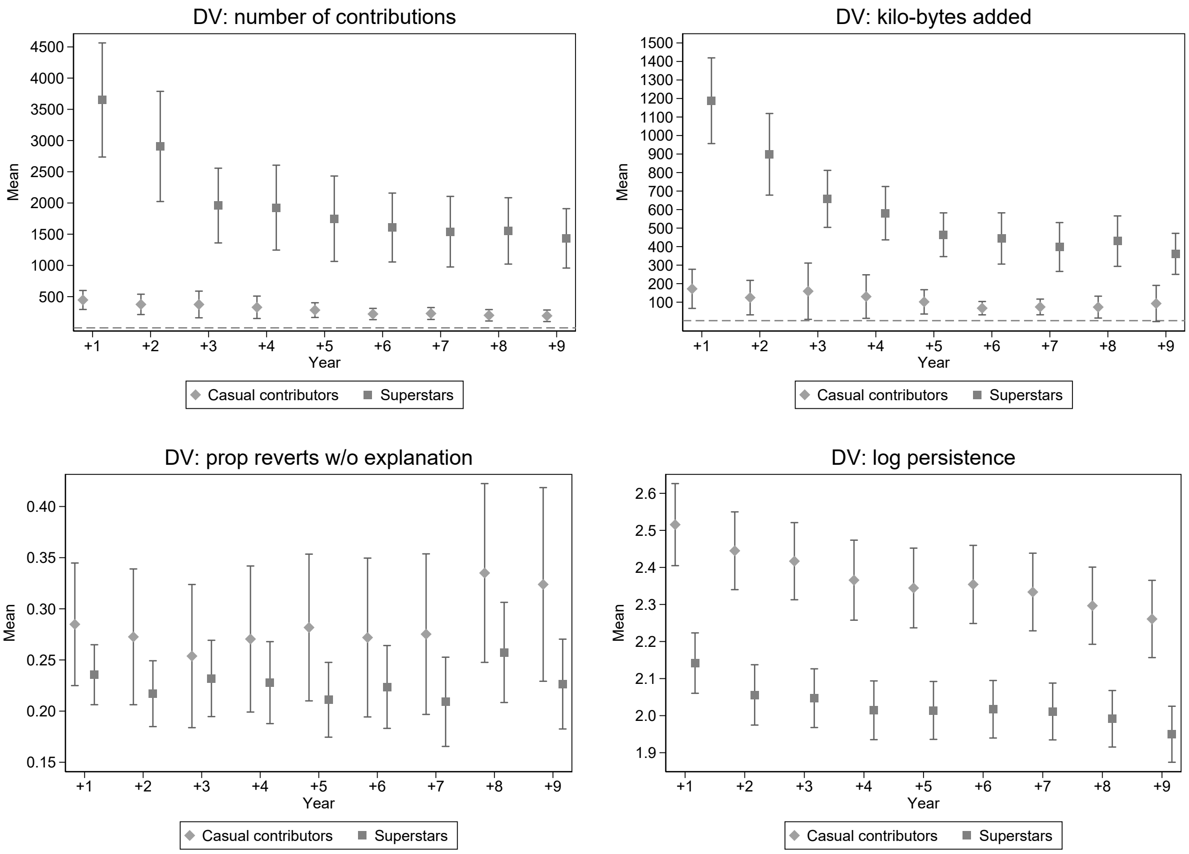


FIGURE (6) Social Motives and Cooperation: Dynamics

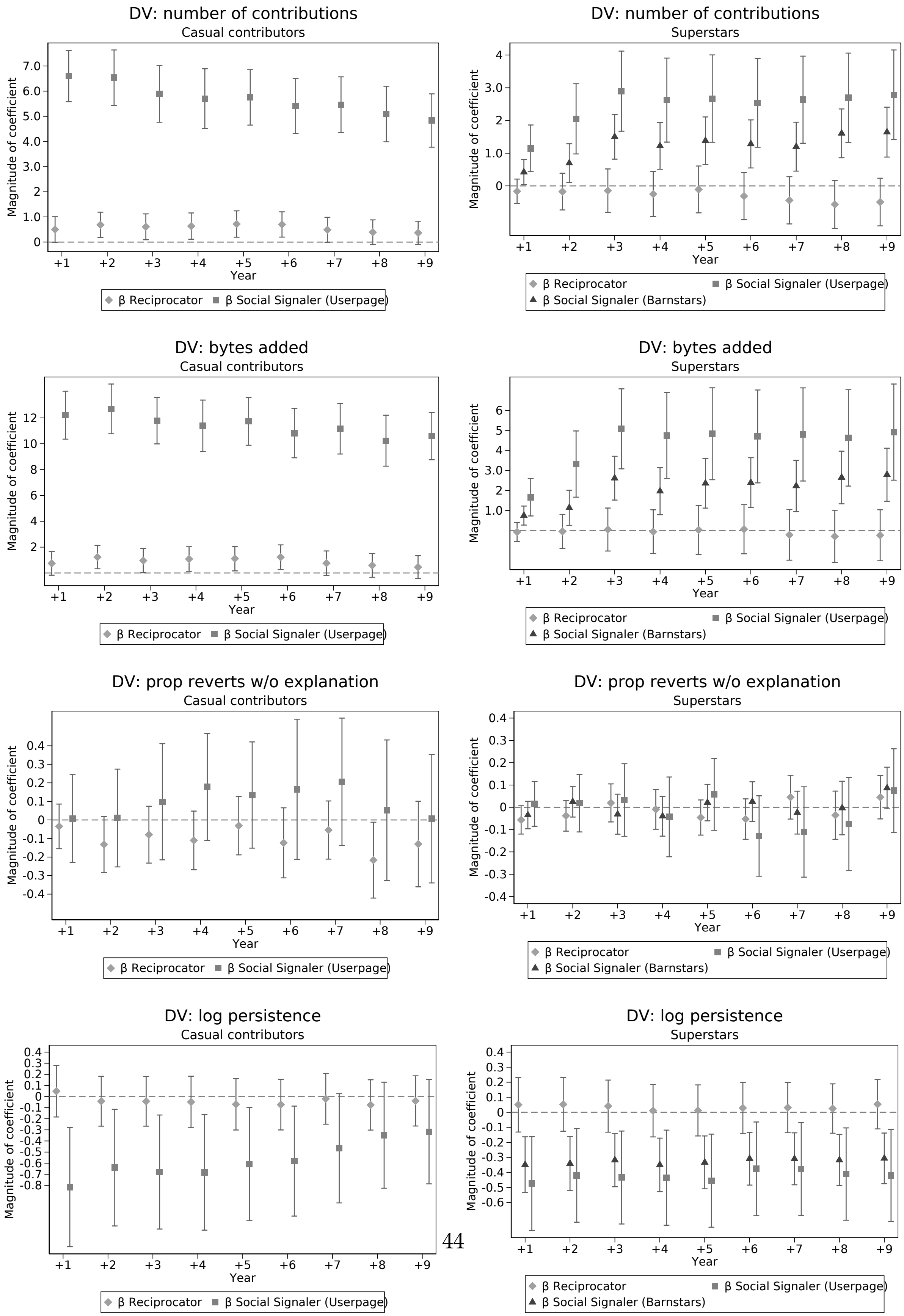


TABLE (1) Common Demographic Characteristics: Wikipedia Editor Survey vs. Present Study

	Wikimedia study	Our study	p-value
Mean age	32.04	32.74	0.207
	(13.99)	(14.36)	
N	5,130	730	
Proportion female	0.09	0.10	0.156
	(0.28)	(0.31)	
N	5,100	730	
Education level	2.88	2.86	0.728
	(1.06)	(1.04)	
N	5,281	722	

Standard deviations in parentheses. Education level is categorized as follows: 1 = "Primary"; 2 = "Secondary"; 3 = "Bachelors"; 4 = "Master's"; 5 = "PhD".

TABLE (2) Sample Descriptive Statistics

	All subjects	Casual contributors	Superstars	p-value
DEPENDENT VARIABLES				
1. Quantity of contributions				
(i) Median number of contributions	889	167	3,222	<0.001
	[0; 21,797]	[0; 5,166]	[295; 41,659]	
(ii) Median bytes added	251,508	33,491	1,280,187	<0.001
	[0; 7,997,347]	[0; 1,015,938]	[65,047; 1.51e+07]	
2. Interpersonal cooperation				
Mean proportion of reverts w/o explanation	0.26	0.29	0.25	0.094
	(0.27)	(0.31)	(0.24)	
3. Quality of contributions				
Mean average log persistence of contributions	2.09	2.26	1.95	<0.001
	(0.80)	(0.88)	(0.70)	
SOCIAL PREFERENCES				
Public goods game data				
- Proportion of altruists	0.09	0.07	0.11	0.034
- Proportion of reciprocators	0.38	0.37	0.38	0.728
- Proportion of weak reciprocators	0.47	0.48	0.45	0.366
- Proportion of free-riders	0.07	0.08	0.06	0.218
Social rewards and signaling data				
Proportion of Barnstar receivers (“Superstars”)	47.50	0	100	
Mean number of Barnstars received	.	0	6.08	
	(.)	(0)	(8.51)	
Size Userpage (in bytes)	4449.33	1793.93	7380.22	<0.001
	(367.78)	(280.16)	(675.70)	
- Social signaler score (from Userpage rating task)	0.34	0.19	0.51	<0.001
	(0.30)	(0.22)	(0.28)	
- Proportion of Superstars signaling Barnstars (from field)	.	0	0.50	
CONTROL VARIABLES				
Age	32.74	30.75	35.04	<0.001
	(14.36)	(13.11)	(15.32)	
Proportion female	0.10	0.10	0.11	0.486
	(0.31)	(0.30)	(0.32)	
Degree level	4.43	4.38	4.49	0.423
	(1.84)	(1.89)	(1.78)	
Salary level	3.68	3.57	3.80	0.200
	(2.31)	(2.33)	(2.29)	
Risk aversion	5.74	5.73	5.75	0.919
	(2.35)	(2.34)	(2.36)	
Total number of subjects	730	383	347	

For the quantity of contributions variables in the top panel: the 10th and 90th percentiles of the distribution are reported in brackets, and the p-values are from Wilcoxon-Mann-Whitney tests of equality of distribution between casual contributors and superstars. For all other variables: standard deviations are reported in parentheses and the p-values are from t-tests. The social signaler score is the mean of the scores received by subjects in the social image rating task (it ranges between 0 and 1; rescaled from a 0 to 10 ranking). Degree level: 1 = “less than high school”; 2 = “high school”; 3 = “some college”; 4 = “2 years college degree”; 5 = “4 years college degree (BA, BS)”; 6 = “masters degree"; 7 = “professional degree (MD, JD)”; 8 = “doctoral degree”. Salary range (monthly): 1 = “0 USD”; 2 = “less than 1000 USD”; 3 = “between 1000 and 2000 USD”; 4 = “between 2000 and 3000 USD”; 5 = “between 3000 and 4000 USD”; 6 = “between 4000 and 5000 USD”; 7 = “between 5000 and 7500 USD”; 8 = “between 7500 and 10000 USD”; 9 = “more than 10000 USD”. Risk aversion: from 0 = “unwilling to take risks” to 10 = “fully prepared to take risks”.

TABLE (3) Pairwise Correlation between the Dependent Variables

	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
Ln(1 + nb contrib.)	1			
Ln(1 + bytes added)	0.957	1		
	($p < 0.001$)			
Prop(rv w/o expl.)	-0.089	-0.088	1	
	($p = 0.039$)	($p = 0.042$)		
Avg(log persist.)	-0.203	-0.278	-0.173	1
	($p < 0.001$)	($p < 0.001$)	($p < 0.001$)	

TABLE (4) Public Goods Contributions: Wikipedia Contributors vs. Standard Subjects Pool

	(1)	(2)	(3)	(4)	(5)
	prop(condit_contrib)	free-rider	Weak reciprocator	Reciprocator	Altruist
Superstar	0.0384*	-0.0143	-0.0397	0.0192	0.0252
	(0.0196)	(0.0172)	(0.0400)	(0.0381)	(0.0183)
Student subject	-0.116***	0.0391	0.0892	-0.0905*	-0.0460**
	(0.0271)	(0.0346)	(0.0579)	(0.0523)	(0.0212)
Controls	Yes	Yes	Yes	Yes	Yes
N	798	798	798	798	798
adj. R^2	0.08	.	.	.	.

The table presents OLS estimates (column (1)) and probit marginal effects (columns (2)-(5)) with robust standard errors in parentheses (constant not reported). The reference group is that of casual Wikipedia contributors. All regressions include the set of individual control variables reported in Table 2, i.e., subjects' age, gender, degree level, salary level and risk aversion. ***, ** and * denote statistical significance at the p <0.01, p <0.05 and p <0.1 levels, respectively.

TABLE (5) Social Motives and Cooperation on Wikipedia

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
Reciprocator	0.454**	0.675**	-0.0465*	0.00702	0.788***	1.012**	-0.0634	-0.0477
	(0.207)	(0.322)	(0.0245)	(0.0697)	(0.299)	(0.495)	(0.0434)	(0.114)
Reciprocator × Superstar					-0.955**	-1.126**	0.0288	0.100
					(0.375)	(0.561)	(0.0522)	(0.141)
Altruist	0.566	1.129**	-0.0605	-0.104	0.105	0.535	0.0204	-0.0513
	(0.365)	(0.529)	(0.0378)	(0.124)	(0.447)	(0.765)	(0.0856)	(0.222)
Altruist × Superstar					-0.190	-0.523	-0.106	-0.0419
					(0.552)	(0.832)	(0.0939)	(0.263)
Signaler (Userpage)	6.160***	9.445***	0.00267	-0.480***	8.801***	14.20***	0.0103	-0.321
	(0.396)	(0.601)	(0.0415)	(0.135)	(0.545)	(0.915)	(0.0839)	(0.234)
Signaler (Userpage) × Superstar					-7.132***	-12.19***	-0.00116	-0.0865
					(0.687)	(1.050)	(0.0947)	(0.276)
Superstar					4.378***	6.983***	-0.0241	-0.183
					(0.359)	(0.520)	(0.0533)	(0.148)
Age	0.0327***	0.0409***	-0.00349***	-0.00316	0.0302***	0.0375***	-0.00342***	-0.00279
	(0.00829)	(0.0126)	(0.000809)	(0.00298)	(0.00753)	(0.0111)	(0.000803)	(0.00297)
Female	-0.969***	-1.431***	0.0147	0.0171	-1.073***	-1.604***	0.0149	0.0264
	(0.317)	(0.534)	(0.0425)	(0.119)	(0.263)	(0.441)	(0.0436)	(0.120)
Degree level	0.0415	0.131	-0.0126	0.0588**	0.00290	0.0640	-0.0135	0.0573**
	(0.0627)	(0.101)	(0.00837)	(0.0241)	(0.0559)	(0.0894)	(0.00841)	(0.0243)
Salary level	-0.000612	-0.0171	0.00655	0.0128	0.00369	-0.00783	0.00716	0.0130
	(0.0494)	(0.0796)	(0.00537)	(0.0166)	(0.0424)	(0.0669)	(0.00537)	(0.0166)
Risk aversion	-0.0836**	-0.102	-0.00237	-0.0103	-0.0672*	-0.0720	-0.00205	-0.0104
	(0.0405)	(0.0637)	(0.00514)	(0.0153)	(0.0356)	(0.0560)	(0.00515)	(0.0153)
Nb Barnstars	0.0733***	0.0827***	-0.00322***	-0.00760	0.0630***	0.0718***	-0.00276**	-0.00385
	(0.0193)	(0.0274)	(0.00116)	(0.00479)	(0.0151)	(0.0198)	(0.00119)	(0.00448)
Size Userpage (in bytes)	-6.69e-06	-1.52e-05	1.11e-06	8.68e-06***	2.22e-06	-5.76e-07	1.08e-06	8.50e-06***
	(9.19e-06)	(1.38e-05)	(1.07e-06)	(3.22e-06)	(7.80e-06)	(1.00e-05)	(1.09e-06)	(3.25e-06)
N	647	647	477	533	647	647	477	533
adj. R^2	0.451	0.419	0.0481	0.0376	0.569	0.555	0.0455	0.0405

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

TABLE (6) The Heterogeneity of Motives: Wikipedia's Superstars

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Prop(rv w/o expl.)	Avg(log persist.)	Avg(log persist.)
	Quantity 1	*Quantity 1*	*Quantity 2*	*Quantity 2*	*Interactions*	*Interactions*	*Quality*	*Quality*
Reciprocator	-0.157	-0.187	-0.0947	-0.133	-0.0348	-0.0345	0.0535	0.0636
	(0.224)	(0.229)	(0.261)	(0.268)	(0.0290)	(0.0292)	(0.0837)	(0.0822)
Altruist	-0.0284	-0.180	0.147	-0.0354	-0.0902**	-0.0923**	-0.126	-0.0955
	(0.338)	(0.326)	(0.358)	(0.334)	(0.0382)	(0.0388)	(0.145)	(0.143)
Signaler (Userpage)	1.707***		2.140***		-0.00354		-0.421***	
	(0.422)		(0.510)		(0.0469)		(0.156)	
Signaler (Barnstar)		0.692***		0.937***		-0.0203		-0.306***
		(0.230)		(0.263)		(0.0281)		(0.0859)
Age	0.0279***	0.0305***	0.0245***	0.0278***	-0.00291***	-0.00292***	-8.23e-05	-0.000823
	(0.00842)	(0.00872)	(0.00848)	(0.00871)	(0.000923)	(0.000925)	(0.00343)	(0.00338)
Female	-0.631*	-0.727*	-0.769	-0.885*	0.0108	0.0105	-0.00665	0.0112
	(0.348)	(0.371)	(0.501)	(0.533)	(0.0400)	(0.0402)	(0.111)	(0.112)
Degree level	-0.0239	-0.0388	0.0153	-0.00340	-0.0112	-0.0112	0.0645**	0.0699**
	(0.0682)	(0.0672)	(0.0788)	(0.0766)	(0.01000)	(0.00996)	(0.0279)	(0.0279)
Salary level	-0.0362	-0.0261	-0.0303	-0.0171	0.00457	0.00422	0.00915	0.00363
	(0.0509)	(0.0536)	(0.0547)	(0.0575)	(0.00622)	(0.00619)	(0.0197)	(0.0201)
Risk aversion	-0.0121	-0.00861	0.0174	0.0213	0.000567	0.000731	0.00740	0.00774
	(0.0424)	(0.0426)	(0.0509)	(0.0510)	(0.00612)	(0.00614)	(0.0166)	(0.0165)
Nb Barnstars	0.0630***	0.0584***	0.0726***	0.0654***	-0.00313**	-0.00273**	-0.00361	0.000248
	(0.0145)	(0.0153)	(0.0185)	(0.0190)	(0.00122)	(0.00120)	(0.00442)	(0.00428)
Size Userpage (in bytes)	9.71e-07	1.14e-05	-4.26e-06	8.65e-06	1.92e-06*	1.92e-06*	7.88e-06**	5.67e-06*
	(9.36e-06)	(8.97e-06)	(1.17e-05)	(1.04e-05)	(1.14e-06)	(1.10e-06)	(3.73e-06)	(3.28e-06)
N	308	308	308	308	292	292	292	292
adj. R^2	0.204	0.183	0.198	0.178	0.0435	0.0452	0.0356	0.0555

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p<0.01$, $p<0.05$ and $p<0.1$ levels, respectively.

APPENDIX

A Related Literature

A.1 Related lab-in-the-field literature

Our paper adds to the growing "lab-in-the-field" literature: a stream of experimental research that does *not* rely on field "interventions" to estimate treatment effects (Gneezy and Imas, 2017; Levine et al., 2023). Instead, lab-in-the-field designs rely upon validated experimental paradigms to elicit (otherwise difficult to observe) underlying preferences in a population of theoretical interest, in order to study how these preferences drive naturally-occurring field outcomes. In this respect, lab-in-the-field designs arbitrate between "identification" and "ecological validity". Field experiments identify cleaner causal effects, but only within the sub-population of "compliers" with the treatment. They also deliver estimates that vary depending on the details on the intervention. By contrast, lab-in-the-field estimates seek to characterize the ecological behavior of the underlying population of theoretical interest, thanks to tighter experimenter control, a cleaner elicitation of subjects' preferences and motives, and the possibility to make direct comparisons across contexts and populations.

In a seminal lab-in-the-field paper, Karlan (2005) thus conducted a Trust game in a population of borrower from a Peruvian microcredit program and found that debt repayment rates depended on reciprocal preferences. Similarly, Barr and Serneels (2009) conducted a Trust game among firm workers in Ghana to uncover a relationship between reciprocal behavior and aggregate labor productivity. Carpenter and Seki (2011) conducted a repeated Public Goods game among Japanese fishermen and showed that more reciprocating fishing crews are more productive. Fehr and Leibbrandt (2011), Leibbrandt (2012), and Gneezy et al. (2016) conducted a battery of social preferences games among Brazilian shrimp catchers and fishermen. They notably showed that prosociality relates to field behavior, such as the propensity to engage in overextraction of resources. Rustagi et al. (2010) ran social dilemma games in the context of small forest communities in Ethiopia. They showed that groups with a larger share of conditional cooperators are more successful in forest commons management, as measured by the quantity of crop trees. Finally, in a lab-in-the-field study of altruistic and social image motives within a population of volunteer firefighters in their communities, Carpenter and Myers (2010) showed that both altruism and signaling motives relate to the decision to join the volunteer fire service.

In a more refined lab-in-the-field design, Kosfeld and Rustagi (2015) identified the social type of community leaders to show that their profile relates to the success of local group in managing forest commons.

Leaders who emphasize equality and efficiency are associated with more positive forest outcomes than anti-social ones. Hergueux et al. (2023) put some bounds on the positive group outcomes associated with reciprocal preferences by conducting a lab-in-the-field study involving *both* alive and failed virtual teams. They showed that highly reciprocal teams are not more successful on average: reciprocal preferences reinforce the cooperative equilibrium in good times, but also make it harder to recover from negative contribution shocks (the project dies).

A.2 Related literature on Wikipedia

Our study adds to a large literature in computer and information science which has long sought to describe the functioning of Wikipedia as an online volunteering or "peer production" community (Benkler, 2002; Benkler et al., 2015). In a seminal ethnographic study of Wikipedia contributors, Bryant et al. (2005) conceptually distinguished regular editors (whom they called "peripheral" or "novice") from superstars (whom they called "experts" or "Wikipedians"). Their qualitative interviews revealed that the most distinguishing feature of the latter group was its high level of self-identification with Wikipedia's goal and community. In a related study, Kittur et al. (2007) provided one of the earliest comprehensive field accounts of Wikipedia editing based on observational data. They showed that compared to superstar (or "elite") editors, regular (or "common") editors accounted for about 70% of overall content creation on Wikipedia. In a follow-on quantitative study, Panciera et al. (2009) replicated those results, and studied the editing trajectories of superstar contributors. Their dynamic analysis revealed a striking pattern: compared to regular users, superstar editors display (and sustain) extreme levels of activity from the very start of their participation. They conclude that "Wikipedians are born, not made."

Relatedly, our paper contributes to a growing literature in economics that studies cooperation in Wikipedia (e.g., Greenstein and Zhu (2018); Greenstein et al. (2021)) and, in particular, the motives that drive field contributing behavior. In a seminal paper, Zhang and Zhu (2011) exploited a natural experiment at Chinese Wikipedia to investigate the role of group size on incentives to contribute. They showed that an exogenous reduction in the size of the community of contributors led to a *decrease* in individual contributions among remaining ones. The authors thus hypothesize that, above and beyond prosocial preferences (i.e., reciprocity or altruism), significant "social benefits" may accrue to some contributors as the size of their community grows, i.e., those editors might be image-driven. Aaltonen and Seiler (2016) used observational data from Wikipedia to show that there is path-dependency in the number contributions received by an article, which is consistent with reciprocal (or "social exchange") dynamics driving contributions at the article level. In a similar vein, Hinnosaar et al. (2022) ran a field experiment which randomly added content to some Wikipedia articles while leaving others unchanged. They found that the treatment increased editing activity on the target articles, but less significantly so than what Aaltonen and Seiler (2016) suggested.

In two consecutive field experiments, Restivo and Van De Rijt (2012) and Restivo and van de Rijt (2014) studied the causal impact of social rewards on contributing behavior. To do so, they randomly allocated community awards (i.e., Barnstars) to a sample of Wikipedia editors taken from the top 10% in terms of their number of edits. They found that receiving a community award causes an increase in the number of edits made over the next three months – but only within the top 1% of editors in terms of activity. When compared to the control group, receiving the award had no effect on the activity of remaining editors, and even *decreased* their retention rate. In a similar field experiment ran on German Wikipedia, Gallus (2017) randomly allocated Barnstars to *newly registered editors* who had made at least two contributions in their first month of activity. She found that receiving a community award causes an increase in the retention rate of those novice editors, namely, the probability that they make at least one edit in the subsequent month increases from 35% in the control group, to 42% in the treatment.

The above field experiments provide clear causal evidence that social rewards impact field contributions, which is consistent with the idea that the behavior of (at least some) contributors is driven by image concerns (social or self). There are important limits to the ecological validity of those field results, however. First, while the typical social practice on Wikipedia is to detail the contribution(s) for which the recipient is being awarded a Barnstar (Kriplean et al., 2008; McDonald et al., 2011; Sajnani et al., 2011), those field interventions could only use some generic text expressing general community appreciation for their contributions. Second, the awards were posted from a special purpose Wikipedia account with no prior editing history, which may have affected the meaning attached to them. Third, the samples on which those causal estimates were obtained are highly selected. In the case of Restivo and Van De Rijt (2012) and Restivo and van de Rijt (2014), the intervention focused on a randomly selected subset of Wikipedia's most active editors *who had never received any community award* so far. In the case of Gallus (2017), the intervention focused on a sample of novice editors, a population which is typically not destined to receive such awards (and may thus have had difficulties comprehending its meaning).

By contrast to those field experiments, our design does not seek to randomly manipulate subsets of the population of contributors to estimate some average treatment effect, resulting from the aggregation of heterogeneous individual preferences. The reason for this design choice is that we are directly interested in the elicitation of the heterogeneity in social preferences within our population of theoretical interest, to see how these may explain the variability in *naturally-occurring* field outcomes.

B Experimental Procedures for the Online Public Goods Game

Conducting a power analysis to determine *ex ante* the optimal sample size for statistical inference is a notoriously uncertain exercise in lab-in-the-field contexts. That said, our initial design was aiming at a sample size about twice as large as the one we eventually assembled. Unfortunately, our subjects recruitment process got interrupted halfway through by the unilateral move of a Wikipedia community administrator who, as he misunderstood the functioning of the banner, decided to temporarily take it down. This unexpected event introduced delays and calendar considerations with respect to the further use of the banner space for this research, as the "official" Wikipedia editor survey was planned to launch right after our experiment, followed by the annual Wikimedia fundraising. We therefore decided, both for logistical and validity reasons, to end our recruitment process there.

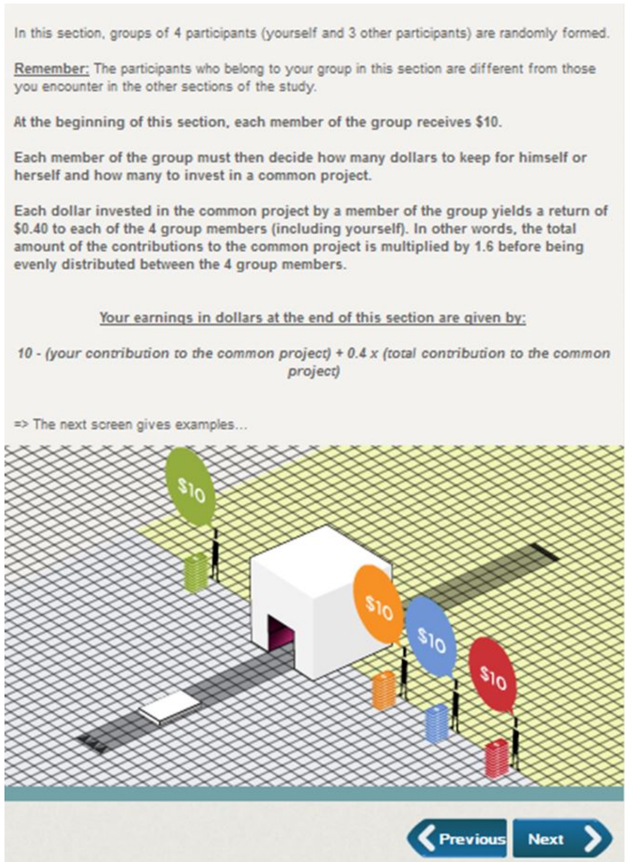
Our online experiment relied on a fully self-contained interface designed specifically to increase the reliability of the experimental data collected over the Internet (Hergueux and Jacquemet, 2015). The welcome page of the decision interface provided subjects with general information about the experiment, including the number of sections, expected completion time, and how earnings are computed. Specifically, in the public goods game, for a player i making a contribution $contrib_i$, the final private payoff is given by:

$$\pi_i = 10 - contrib_i + 0.4 \sum_{j=1}^{4} contrib_j. \tag{3}$$

In order to minimize potential demand effects and in-group biases, we were very careful not to present the study as being Wikipedia oriented. We made it very clear on the introductory screen that subjects would interact with a diverse pool of Internet users. Final earnings were computed by randomly matching our subjects with individuals from a pool made up of open source software developers and students.

One important methodological aspect of the online implementation of the experiment is to guarantee a quick and thorough understanding of the instructions when no interaction with the experimenter is possible. We strengthened the internal validity of our online experiment through three distinctive features of the interface. First, we included novel flash animations illustrating the written experimental instructions at the bottom of the instruction screen (see Figure A1). Second, the instruction screen was followed by a screen providing some examples of decisions, along with a detailed calculation of the resulting payoffs for each player. These examples were supplemented on the subsequent screen by an earnings calculator. On this interactive page, subjects were allowed to test any scenario they wanted to consider. Finally, the system provided quick access to the instruction material at any moment during decision-making.

FIGURE (A1) The instruction screen of the Public Goods game



At the end of the experiment, subjects' final payoffs were added to their $10 participation fee. Payments were made via an automated PayPal transfer. Such a payment procedure guarantees a fungibility similar to that of cash transfers in lab experiments since money transferred via PayPal can be readily used for online purchases or easily transferred to one's personal bank account at no cost. We only required a valid e-mail address to process the payment. To strengthen the credibility of the payment procedure, we asked subjects to enter the e-mail address that was (or would be) associated with their PayPal account right after the introductory screen of the decision interface.

It is important to stress that Wikipedia contributors can be very hostile to monetary rewards. In order to ensure that the experiment was equally incentive-compatible for all subjects, we allowed them to donate their final earnings to the Wikimedia Foundation and/or the International Committee of the Red Cross upon completion of the experiment. This possibility was made clear on the welcome screen of the decision interface.

C Wikipedia user pages and the social signaling rating task

The possibility to create a personal user page on Wikipedia is first made apparent upon registering an account, but creating a Wikipedia user page is possible at any moment. In practice, the page is created by virtue of its owner posting some content to it. Because user pages are personal, their content vary greatly in size and nature. Wikipedia guidelines only prevent editors from using this page as a weblog or personal website, or from posting extensive material unrelated to Wikipedia. Some editors thus organize their user pages as small Wikipedia biographies, describing their personal trajectories as contributors and providing personal thoughts on the functioning of the community. Others use it as a notebook that publicly details their past and intended future contributions. Finally, some decide to create their user page, but purposefully leave it blank.

Evaluating a Wikipedia user page is not necessarily easy for an uneducated eye, as comprehending its content often requires some level of insider knowledge. We thus developed a set of instructions aimed at introducing our online raters to Wikipedia's main social signaling practices. First, editors largely communicate about themselves on their user page through different kinds of "userboxes": small stickers that can be personalized at will to signal any user-specific quality or interest. Many editors use this tool to signal their hobbies, the languages they speak, the countries they have visited, or even their socio-political views to other community members. But many also use it for more direct social signaling purposes, such as keeping public score of their edit count, or signaling their seniority within the community of contributors. Beyond the use of userboxes, editors also frequently seek to advertise their work by maintaining personal tables or lists of their contributions and edited articles. Last but not least, many editors seek to signal their skills, achievements and social standing to the broader community by presenting, in a separate section of their user page, a "gallery" of Barnstars received from other editors.

We run our social signaling rating task on the *Prolific.co* platform. Similar to *Amazon Mechanical Turk*, *Prolific.co* provides researchers with the opportunity to distribute tasks that require human judgement (or even run experiments) over a large population of online workers (Horton et al., 2011; Paolacci et al., 2010; Buhrmester et al., 2011; Thomas and Clifford, 2017). We favored the *Prolific.co* platform because it collects the individual identities of its workers, together with some basic demographic information. Those additional features make its overall pool of workers more reliable, allow to screen study participants based on predetermined factors, and generate significantly higher quality data (Peer et al., 2017; Chmielewski and Kucker, 2020; Peer et al., 2022; Douglas et al., 2023).

We further increased the internal validity of our data by adding two features to our design. First, we leveraged recent experimental research on the positive effect of oath-taking on honesty and cooperativeness by requesting workers to take a solemn oath of honesty at the very start of the study (Jacquemet et al., 2013, 2019, 2021; Hergueux et al., 2022). To proceed, workers thus had to agree to "swear upon their honor" that they would "read the instructions carefully and always provide honest answers." Second, we incentivized effort within our sample of workers by offering an additional £5 bonus payment to the 10% of participants "whose ratings are most consistent with those obtained from an independent expert (who previously performed the same task)."[26] For those "high-quality" raters, this bonus therefore amounted to a 111% increase in individual earnings (corresponding to an hourly wage of about $24.20).

D Alternative Measurement Strategies for our Dependent Variables

D.1 Interpersonal cooperation

In this paper, we focus on subjects' reverting behavior to proxy for their level of interpersonal cooperativeness. By doing so, we purposefully exclude two other variables that have also been used for related purposes, albeit somewhat differently, in the Wikipedia context. The main reason for this choice is that reverting one's work without justification clearly represents a non-cooperative act. As the literature reported in the main text shows, reverting editors often pay a small messaging cost relative to the collective benefits reaped from the resulting increase in editor motivation and retention.

By comparison, a related stream of Wikipedia research has focused on "edits wars" as another important instance of low quality interactions while editing. An edit war is a sequence of reverts where editors override each other's contributions in turn. Such behavior is typically the result of an editing conflict between editors, and is often considered harmful, as editors are encouraged to resolve disagreements through discussions.[27] The literature has thus developed tools aimed at identifying sequences of consecutive reverts to measure and document article-level conflictuality Yasseri et al. (2012, 2014); Chhabra et al. (2020), or study how exposure to "extreme pushback" from another editor affects users' preexisting slant or bias (Greenstein et al., 2021). Remarkably, while this literature readily infers "conflictuality" and "extreme pushback" from edit war events, it refrains from directly interpreting them in terms of editor cooperativeness. To be sure, some editors engage in edits wars in good faith, simply by virtue of trying to protect the common resource against a troll or vandal. As a result, interpreting edit war situations in terms of editor cooperativeness requires a more detailed contextual examination of editors' field interactions, in order to assess their respective intentions. Relatedly, sentiment analysis has been used, e.g., to describe systematic differences in the communication patterns of women editors and Wikipedia administrators

[26] In practice, we used the user page ratings we obtained from our testing phase as the standard against which we awarded those bonuses.

[27] See https://en.wikipedia.org/wiki/Wikipedia:Edit warring

(Laniado et al., 2012). However, negative sentiment does not necessarily reflect low cooperativeness. For instance, empirical evidence suggests that leaders of online volunteering communities equally resort to positive and negative sentiment in their communication Hergueux and Kessler (2022).

D.2 Quality of contributions

Related to the empirical strategy we report in the main text, a more recent line of research has taken a machine learning approach to estimate the quality of editors' contributions (Halfaker and Geiger, 2020). For instance, Li et al. (2020) relied on the Wikimedia ORES classifier – designed to predict the stage of development of a given article – to compare the probabilities assigned to various development stages over two consecutive versions. They thus derive an indicator of edit quality. This approach is not without problems. While individual edits vary greatly in size, this method takes the edit as its smallest unit of analysis. Further, it assumes that the article development scale is an interval one, although the marginal cost of moving articles up the latest development stages is significantly higher (Viégas et al., 2007). Even more recently, Wikipedia classifiers have evolved to address those limitations and make direct edit quality predictions, notably by estimating the probability that an edit was made in "good faith", or is "damaging" (TeBlunthuis et al., 2021).[28] In essence, those classifiers rely on community based labels to generate their predictions. There is therefore no reason *a priori* to believe that such quality measures may systematically differ from those we derive from editors' peer reviewing work. This, however, is an open question for future research.

E Alternative Specifications for Main Results

E.1 Without control variables

In this Section, we rerun our baseline specification from Table 5, but exclude our set of individual level control variables together with *Nb_Barnstars* and *Size_Userpage*. Restricting the set of independent variables to our three social preferences of interest can be informative in two (complementary) ways. First, it allows for a direct comparison of the magnitude and statistical significance of our estimates with and without control variables, which provides information on the magnitude of any (controlled for) omitted variables bias. Second, it allows to compute informative tests of the possible range of *unobserved* omitted variables bias – expressed in terms of the overall correlation between the variables of interest and the

[28]See https://www.mediawiki.org/wiki/ORES, as well as the most recent version of this machine learning classifier: https://wikitech.wikimedia.org/wiki/Machine_Learning/LiftWing

observables in the full model – which preserves the statistical significance of our estimates (Oster, 2019).

We present our results in Table A1. Comparing these coefficients to those from Table 5 across columns (1)-(8), we can see that our estimates remain quite stable between specifications. The strongest point-estimate deviations we observe relate to the coefficients on altruistic preferences in columns (1)-(3), where omitted variables bias appears to drive up the magnitude of these coefficients, so that they reach statistical significance in this restricted model. By contrast, excluding our control variables from the model leaves our point-estimates on reciprocal and social image preferences relatively less affected. However, doing so actually *reduces* the statistical significance of the relationship between reciprocal preferences and contribution quantity (columns (1)-(2)), as a result of the increased variance left unexplained in the error term. Interestingly, interacting our model with subjects' superstar status appears to substantially reduce the scope for omitted variables bias, as can be seen from comparing columns (5)-(8) in Tables 5 and A1.

As a second step, we perform a sensitivity analysis of our model to possible *unobserved* omitted variables bias (Oster, 2019). Given our full set of control variables, the idea behind this analysis is to estimate the minimum strength, or "breakdown point" b, of the correlation between the variable of interest and a (hypothetical) set of unobservables, which would result in the associated coefficient losing its statistical significance. The breakdown point b is expressed in terms of the observed correlation between the variable of interest and the *actual* set of controls. In other words, given our observed set of control variables, the causal effect of the variable of interest (that is, in the absence of omitted variables), will remain statistically different from 0 as long as the correlation between this variable and the unobservables is at most b% as large as its correlation with the observables. The result of this thought experiment obviously depends on the nature and quality of the available set of controls, but is nonetheless informative in a given empirical context.

We limit our sensitivity analysis to the average (i.e., non interacted) models reported in columns (1)-(4) of Table 5, both because these average coefficients appeared less stable than those reported in models (5)-(8), which should deliver a more stringent criterion, and in order not to spread our discussion over too many breakdown points. We find that, given our observed set of control variables, the breakdown point b associated with our coefficients on reciprocal preferences is of 30.2 and 28.9% in models (1)-(2), 28% in model (3), and 1.5% in model (4). Similarly, with respect to altruistic preferences, we find b=18.4 and 23.4% in models (1)-(2), b=18.5% in model (3), and b=10.5% in model (4). Finally, with respect to social image preferences, we find b=66.5 and 55.2% in models (1)-(2), b=0.5% in model (3), and b=27% in model (4). Across the range of statistically significant coefficients reported in Table 5, our above analysis therefore suggests that the ones that may be most adversely affected by potential unobserved omitted variables bias would be that on altruistic preferences, while the most resilient ones would be that on the average relationship between social image preferences and contribution quantity.

E.2 Same N across regressions

In our baseline results from Table 5, the number of observation per column varies depending on whether we could measure our target dependent and control variables for each subject. We document in Sections 2.4.2 and 2.4.3 that, contrary to our contribution quantity variables which are, by definition, available for all subjects, we could only compute our measure of cooperativeness while editing for 538 out of 730 subjects, and our measure of contribution quality for 605 out of 730 subjects. In both instances, a substantial fraction of these cases involve subjects who made no edits over our entire time period (N=84), or subjects with very low edit counts (who, e.g., never reverted another editor, saw their contributions deleted from the database for some reason, or contributed to a Wikipedia page that got subsequently deleted). In addition, because the history logs of the user page of three of our subjects had been deleted by the time we ran our distributed social signaling rating task (see Section 2.3.1), we could only compute our related social image preferences variable for 727 out of 730 subjects. Finally, a number of subjects refused to answer some of our key control variables. Specifically, out of our 730 subjects, 8 refused to answer on their degree level, 72 on their salary level, and 9 on their risk aversion level.

By contrast to Section E.1, where we rerun our baseline specification without including any control variable, in this Section, we rerun our baseline specification from Table 5 with our full set of control variables, but restrict the sample across regressions so that it remains stable across columns. Combined, the missing variables on both our dependent and independent variables restrict our available sample to 457 subjects. In particular, this procedure excludes many subjects with relatively low edit counts (if any) over our time period. However, this exercise is informative in that it allows to compare the obtained estimates when maintaining the subjects pool fixed across columns.

We present the results of this exercise in Table A2. As could be expected, the relationships most affected by these sample restrictions is that of reciprocity and altruism preferences with our quantity of contribution variables (columns (1)-(2)). In both cases, these relationships lose their statistical significance, likely as a result of the reduced variability in the dependent variable. By contrast, across columns (1)-(4), this sample restriction exercise appears to affect our coefficients on social image preferences relatively less. Finally, it is interesting to note that across columns (5)-(8) of Table A2, where we augment our baseline model by interacting our coefficients of interest on social preferences with subjects' superstar status, the resulting increased precision in the estimated relationship within both groups allows to recover statistically significant coefficients that align with our results from Table 5.

E.3 Free-riders as an independent category

In the baseline specification reported in Equation 2, we group the categories of free-riders and weak reciprocators to form the baseline, excluded group in our regressions of field cooperation on social preferences. This choice is justified by the fact that we are most interested in the field behavior of reciprocators and altruists in the public goods game. Further, free-riders constitute a relatively small proportion of the subjects in our experiment (i.e., 7% overall). Combined with the fact that we are interested in estimating a relationship that potentially differs between the group of casual and superstar editors, our limited sample size may make it difficult for our model to estimate an additional social type coefficient (i.e., that on weak reciprocators) without imposing a prohibitive cost in terms of the statistical power required for our tests.

In this Section, we reproduce the results from the baseline model reported in Table 5, but single-out the category of free-riding editors as our baseline category. We do this in two different ways. First, we conduct this analysis using the strict classification of social types detailed in Section 2.1. Second, we relax this classification to allow for some level of within-subject decision error. Specifically, while our baseline classification requires that, on average, subjects make an average conditional contribution to the public good of $m_i = 0$ in order to be classified as free-riders, we now require $m_i \leq 1$ (out of an endowment of \$10). Conversely, while our baseline classification requires that, on average, subjects make an average conditional contribution to the public good of $m_i = 10$ in order to be classified as altruists, we now require $m_i \geq 9$. In practice, this less stringent classification rule results in the reclassification of some weak reciprocators in our data as free-riders. The overall proportion of free-riders thus increases from 7 to 9.6%. Conversely the proportion of altruists in our data rises from 8.6 to 10.4%.

We present the results of this analysis in Table A3. Focusing first on Panel A, which corresponds to the strict classification of social types used in the main text, we can see that singling out free-riders as the baseline category in our regressions prevents the precise estimation of most of our coefficients of interest, except that on social image preferences. That said, we can see from columns (1)-(4) of Panel B, which implements a less stringent classification of social types, that the reclassification of a number of (almost never contributing) weak reciprocators as free-riders allows our coefficients to regain their statistical significance in a way that is qualitatively consistent with our results from Table 5. However, when interacting this baseline model with subjects' superstar status in columns (5)-(8), most of our estimates become again too imprecise to reliably replicate our findinds from Table 5.

E.4 Alternative sample of Superstars

In the main text, we define superstar contributors as Wikipedia editor who received at least one community award from another editor over their entire contribution history (see Section 2.2) – even if the Barnstar was awarded during the time covered by our study (i.e., after the time at which we elicited our main social preferences of interest, on December 8th, 2011). This implementation choice is different than the one we enacted for our other independent variables of interest, which were measured at the time of our experiment. Its justification lies in the findings of the extant empirical Wikipedia literature, which has found that "superstar" type contributors start to behave as such from the very start of their editing career (Bryant et al., 2005; Panciera et al., 2009)) – an observation which is consistent with the dynamics of our data. These results thus suggest that superstar editors identify very early-on with the goals and community of Wikipedia, so that superstar status might be best defined as a stable rather than an evolving within-subject characteristic.

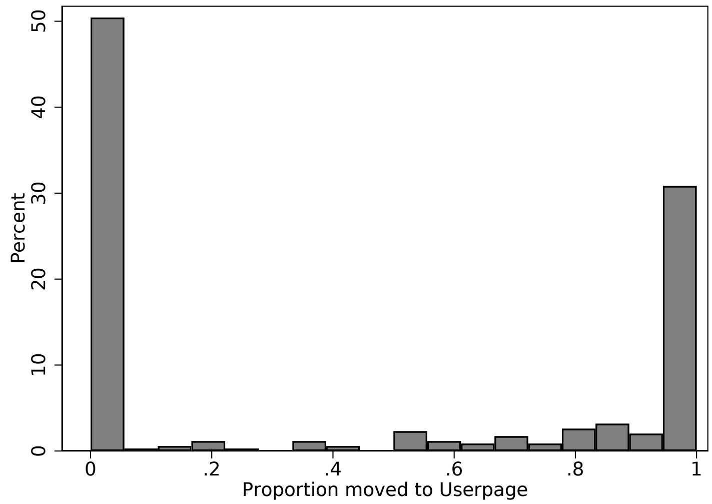
In practice, however, this implementation choice might affect our estimates, as out of the 347 superstar Wikipedia editors in our data, 73 were awarded their first Barnstar after December 10, 2011. The classification of these subjects as "superstars" is therefore subject to interpretation. According to our definition from Section 2.2, those subjects are included in the subgroup of superstar editors.

In this Section, we explore the robustness of our main results from Table 5 to excluding these 73 subjects from our estimations, in order to strictly restrict our sample of superstar editors to those who achieved such status before December 8th, 2011 (i.e., the time at which we ran our experiment). We report our results in Table A4. We can see that, when compared to those from Table 5, our coefficients of interest across columns (1)-(8) remain largely robust to excluding these "late" superstar editors.

E.5 Social signaling as proportion of Barnstars displayed

In the main text, we define Barnstar signaling as an alternative, field-based measure of social image concerns within the group of superstar editors (see Section 2.3.2). Specifically, among subjects who received Barnstars in our sample (the "superstars"), we coded as "social signalers" those who decided to display at least one of their awards on their personal user page. As a potentially superior measure of social image concerns, however, one may have used the *proportion* of subjects' awards that they decided to prominently display on their personal user page. In Figure A2, we thus report the proportion of the awards received by the superstars in our sample that they decided to display on their personal user pages. Two modes largely account for the variation in this data: 50% of superstars never report any Barnstars, while over 30% report them all.

FIGURE (A2) The proportion of Barnstars received by superstar editors displayed on their Userpage



In this Section, we reproduce our analysis in columns (2), (4), (6) and (8) of Table 6 by replacing our baseline field-based measure of social image motives with the above discussed variable. We report our results in Table A5. We can see that, when compared to those from Table 6, our coefficients on social image preferences, if they yield slightly higher point-estimates, remain largely robust to this change in variable definition.

E.6 Controlling for article controversiality

In this Section, we explore the robustness of our main results to one possible confound in our regression analysis: the average controversiality of the set of page edited by our subjects. For instance, it could be that contributors motivated by social image concerns seek social recognition and prestige by focusing their edits on highly controversial topics, which may cause the observed decrease in the overall persistence of their contributions. Such behavior could therefore account for the strong negative relationship we uncover between social image motives and contribution quality: image-driven editors willing to show-off their skills could merely engage with difficult topics that attract significant attention from other (well-recognized) contributors, causing the observed negative association with the mean persistence of their field contributions, but without actually implying a drop in contribution quality.

This section addresses this possibility directly. To do so, we build upon a stream of the empirical literature on Wikipedia which has developed tools aimed at measuring the level of controversiality of

Wikipedia articles. In practice, the idea behind this literature is that controversial topics on Wikipedia can be effectively identified by looking at whether they generate "edit wars" between contributors. An edit war happens on a given article when two contributors engage in a sequence of (at least two) consecutive reverts, thus overriding each other's edits in turn. To be sure, such behavior is typically the result of a strong editing conflict between contributors. A diverse set of reasons can lead contributors to engage in an edit war. However, Wikipedia policies strongly warn editors against such reactions, whatever their underlying justification, as all disagreements need to be eventually resolved through community discussions and consensus building, while edit wars foster antagonism and content instability at the article level.

Based on those observations, Yasseri et al. (2012) developed an algorithm to identify edit wars or "mutual reverts" at the article level. They used this algorithm to measure topic conflictuality on Wikipedia, and study the long term dynamics of conflict resolution. In a related study, Yasseri et al. (2014) relied on the same set of tools to characterize and describe the most controversial topics across 10 different linguistic versions of Wikipedia. Chhabra et al. (2020) built upon the work of Yasseri et al. (2012) by trying to predict the length and outcome of edit wars at the article level based on a set of *a priori* determined features. Greenstein et al. (2021) also focused on mutual reverts as a clear instance of editing conflict, but they were not interested in leveraging this measure to characterize conflictuality at the article level. Rather, they sought to build an *editor level* measure of the number of "extreme pushback" events that they experienced over their editing careers, in order to study its relationship to the evolution of their preexisting ideological slant or bias.

In the context of this paper, we are interested in constructing a control measure of the *average controversiality of the set of articles edited by our subjects*. We rely on the tools developed by the above literature to do so. We first proxy for controversiality at the article level by identifying the number of "mutual reverts" on each article. A mutual revert m is identified on an article when a pair of contributors (j; k) is identified once with j and once with k as the reverter. In other words, we seek to identify controversial topics in Wikipedia by looking at editors' tendency to "undo" each others' work in turn on a given page. We further follow the literature by noting that a mutual revert is more indicative of a strong controversiality potential when it involves two experienced editors than two inexperienced ones (or even one experienced and one inexperienced). We thus weight each mutual revert by the minimum of the overall number of edits N_j and N_k that both contributors made. For each article, we then sum-up the weights of all mutual reverts (we exclude the highest weight in order to avoid overestimating controversiality by giving a lot of prominence to a single editorial conflict). Finally, we multiply this quantity by the total number of individual contributors E who ever got involved in a edit war in the context of the target article, so that articles where mutual reverts involve a higher number of *distinct* editors receive a higher score (Yasseri et al., 2012, 2014). The controversiality score C_a of an article is therefore computed as:

$$C_a = E \times \sum_{m=1}^{n} min[N_j; N_k] \tag{4}$$

The above measure of controversiality is at the Wikipedia article level. In order to construct a controversiality score at the contributor level (e.g., in the spirit of Greenstein et al. (2021)), we compute the average of the controversiality scores over all the articles a that each contributor C_i edited, weighted by the proportion of the overall number of contributions that they made to each article:

$$C_i = \sum_{a=1}^{n} p_a \times C_a \tag{5}$$

Finally, we take the logarithm of the above variable to get a (non-exploding) measure of the average controversiality of the set of articles edited by each subject: *Ln(edited articles controversiality score) = ln*(C_i). This variable has mean = 1.46; std = 1.25; min = 0 and max = 10.42.[29]

The main hypothesis which motivated the present exercise is that, in order to earn social recognition and prestige within the community of contributors, image-driven subjects might focus their contributions on relatively more controversial (and, thus, more "visible") articles. Those focused contributions could then largely account for the lower level of persistence of the content contributed by social signalers on average, which we observe in the data, and generate omitted variables bias on our coefficients of interest. As a first step to this analysis, we therefore start by reporting in Table A6, separately for both subgroups of casual (Panel A) and superstar contributors (Panel B), the pairwise correlation between (i) the average controversiality score of the set of pages edited by our subjects, (ii) the number of Barnstars they received, (iii) their social signaling preferences (as measured by our distributed user page rating task), and (iv) their social signaling preferences (as measured within the group of superstars by looking at whether they chose to prominently display at least one of their awards on their personal page).

Focusing first on Panel A, we can see that, among casual contributors, our measure of social signaling preferences (derived from the online user page rating task), is positively and significantly correlated with the average conflictuality score of the set of pages they edited (corr=0.366, p <0.001). By contrast, turning our attention to Panel B, we see that this is not the case for superstar editors, both according to the above measure of social image preferences, and that based on field Barnstar signaling. In both cases, the correlation coefficient with the average conflictuality score of the set of pages edited by our subjects is

[29]Experimenting with other ways of computing this variable such as taking the absolute value of C_i or the proportion of the articles edited by contributor C_i which have a positive controversiality score C_a leaves our results unchanged.

close to zero and not statistically significant. Further, the number of community awards received by our subjects is not correlated with the average conflictuality score of the articles they edit (this correlation is again close to zero and not statistically significant).

Incidentally, Panel B of Table A6 also shows that our measures of social image concerns are strongly correlated with each other (corr=0.542, p <0.001), and with the overall number of Barnstars received by subjects (corr=0.291, p <0.001, and corr=0.358, p <0.001, respectively). In particular, the fact that superstars who receive a higher number of Barnstars are more likely to prominently display these awards on their personal page (i.e., reveal relatively stronger social image motives), justifies that we include *Nb_Barnstars* as a control in the baseline specification reported in Equation 2, so as to break its correlation with our *Social_signaler* variable of interest (see Section 4.1).

All in all, these correlational results provide mixed evidence in support of the narrative according to which editors motivated to earn social recognition would engage with topics that are more controversial on average. We find evidence consistent with such behavior within the group of casual editors. However, topic controversiality does not correlate with either with the number of community awards received by the Wikipedia superstars, or their social image preferences.

As a second step to our analysis, we directly investigate the possibility that omitting to control for article controversiality in our baseline specification introduces a bias on our estimates of interest. Since we are particularly interested in any possible omitted variables bias on our coefficient on social signaling preferences, we replicate our main results from both Tables 5 and 6 after directly including our indicator of the average controversiality of the set of pages edited by each subject as an additional control variable.

We report our results in Tables A7 and A8, respectively. We can see that, in practice and across both tables, our point-estimates and confidence intervals of interest remain largely unaffected by the inclusion of this additional control variable. Our empirical exercise therefore lands additional support to the interpretation of our results, which we report in the main text: as they seek social recognition, image-driven editors appear to actively favor quantity in their contribution decisions at the expense of quality, as opposed to merely focusing their activity on relatively more controversial topics.

Finally, looking at the estimated coefficients on our controversiality variable across Tables A7 and A8 provides interesting additional insights on editor behavior. First, we can see from Table A7 that, on average, editors who focus their contributions on controversial topics make *more* contributions than others on average (columns (1)-(2) and (5)-(6)), and feature a (marginally significant) decrease in their proportion of unjustified reverts (columns (3) and (7)). However, they do not produce higher quality content on average (columns (4) and (8)). The corresponding results from Table A8, focused on the field behavior of superstar contributors, reveal a striking contrast. In this group, editors who focus their contributions on controversial topics make significantly *less* contributions than others on average (columns (1)-(4)), justify their reverts significantly more often (columns (5)-(6)), and also tend to produce content that persists *more* peer revisions on average (columns (7)-(8)). These results suggest that, by contrast to casual contributors, superstar editors are better aware of the average conflictuality potential of the set of pages that they decide to edit. They therefore preemptively invest in making higher quality, better justified contributions, at the expense of edit quantity. By contrast, casual editors' ability to adjust their contribution strategy as a function of the conflictuality potential of the set of pages that they edit appears significantly lower.

F Altruism and Cooperation: Dynamics

This Section complements the results presented in Section 5 by reporting our disaggregated yearly estimates of the evolution of the relationship between altruistic preferences and field cooperation within both groups of casual and superstar editors. We report these additional results from our main regressions in Figure A3, which mirrors the structure of Figure 6 in the main text. This figure globally supports the conclusion that the coefficients on altruism are rather unstable and not precisely estimated in our data, which makes them more difficult to interpret. The only notably exception relates to the third-right row of the figure, where we report, over our entire time period, a (quite precisely estimated) positive relationship between altruistic preferences and superstars' level of cooperativeness towards others while editing – a result that replicates a significant finding from Table 6.

TABLE (A1) Social Motives and Cooperation on Wikipedia: without control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
Reciprocator	0.374*	0.535*	-0.0520**	0.0469	0.774***	1.006**	-0.0735*	-0.00922
	(0.199)	(0.308)	(0.0242)	(0.0679)	(0.282)	(0.468)	(0.0431)	(0.110)
Reciprocator × Superstar					-0.990***	-1.251**	0.0355	0.0925
					(0.356)	(0.531)	(0.0517)	(0.138)
Altruist	0.958**	1.593***	-0.0952**	-0.0640	0.367	0.848	-0.0421	-0.0385
	(0.375)	(0.529)	(0.0376)	(0.112)	(0.454)	(0.768)	(0.0800)	(0.207)
Altruist × Superstar					-0.0980	-0.468	-0.0642	0.0149
					(0.576)	(0.848)	(0.0899)	(0.245)
Signaler (User page)	6.877***	10.19***	-0.0500	-0.446***	9.094***	14.66***	-0.0607	-0.184
	(0.337)	(0.516)	(0.0373)	(0.116)	(0.506)	(0.859)	(0.0820)	(0.223)
Signaler (User page) × Superstar					-6.795***	-12.06***	0.0327	-0.121
					(0.646)	(0.988)	(0.0932)	(0.262)
Superstar					4.576***	7.367***	-0.0438	-0.228
					(0.337)	(0.493)	(0.0506)	(0.140)
N	727	727	536	602	727	727	536	602
adj. R^2	0.395	0.379	0.0113	0.0219	0.531	0.534	0.00795	0.0330

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

Table (A2) Social Motives and Cooperation on Wikipedia: same N across regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
Reciprocator	0.0583	0.168	-0.0443*	-0.0241	0.496*	0.594*	-0.0719	-0.137
	(0.171)	(0.193)	(0.0253)	(0.0673)	(0.271)	(0.319)	(0.0442)	(0.106)
Reciprocator × Superstar					-0.762**	-0.753*	0.0473	0.193
					(0.352)	(0.400)	(0.0533)	(0.136)
Altruist	0.140	0.306	-0.0440	-0.135	-0.296	-0.339	0.0655	-0.183
	(0.250)	(0.283)	(0.0377)	(0.110)	(0.371)	(0.426)	(0.0904)	(0.217)
Altruist × Superstar					0.360	0.558	-0.144	0.118
					(0.465)	(0.520)	(0.0969)	(0.251)
Signaler (Userpage)	1.740***	2.086***	5.16e-05	-0.426***	1.759***	2.161***	0.0134	-0.253
	(0.338)	(0.374)	(0.0425)	(0.127)	(0.635)	(0.689)	(0.0848)	(0.215)
Signaler (Userpage) × Superstar					-0.577	-0.915	-0.00862	-0.0953
					(0.733)	(0.794)	(0.0965)	(0.264)
Superstar					1.212***	1.639***	-0.0321	-0.252*
					(0.362)	(0.401)	(0.0538)	(0.142)
Age	0.0198***	0.0153**	-0.00358***	-0.00194	0.0174***	0.0121*	-0.00348***	-0.00115
	(0.00659)	(0.00681)	(0.000824)	(0.00302)	(0.00671)	(0.00692)	(0.000814)	(0.00302)
Female	-0.392	-0.447	0.0176	0.0392	-0.472*	-0.571	0.0177	0.0590
	(0.288)	(0.391)	(0.0434)	(0.103)	(0.280)	(0.377)	(0.0448)	(0.0993)
Degree level	-0.0122	0.0178	-0.0126	0.0374*	-0.00340	0.0282	-0.0137	0.0342
	(0.0514)	(0.0637)	(0.00860)	(0.0222)	(0.0514)	(0.0621)	(0.00862)	(0.0223)
Salary level	-0.0269	-0.0226	0.00545	0.0220	-0.0294	-0.0235	0.00630	0.0224
	(0.0408)	(0.0459)	(0.00546)	(0.0158)	(0.0398)	(0.0442)	(0.00548)	(0.0158)
Risk aversion	-0.0307	-0.00920	-0.00425	-0.00941	-0.0344	-0.0150	-0.00385	-0.00904
	(0.0326)	(0.0380)	(0.00534)	(0.0139)	(0.0323)	(0.0372)	(0.00528)	(0.0137)
Nb Barnstars	0.0724***	0.0858***	-0.00359***	-0.00721	0.0605***	0.0686***	-0.00294**	-0.00347
	(0.0147)	(0.0191)	(0.00120)	(0.00469)	(0.0134)	(0.0168)	(0.00120)	(0.00432)
Size Userpage (in bytes)	1.61e-06	-1.63e-06	1.36e-06	7.20e-06**	2.78e-06	2.13e-07	1.37e-06	6.92e-06**
	(7.85e-06)	(9.67e-06)	(1.08e-06)	(2.95e-06)	(7.98e-06)	(9.49e-06)	(1.09e-06)	(3.01e-06)
N	457	457	457	457	457	457	457	457
adj. R^2	0.222	0.224	0.0517	0.0335	0.252	0.272	0.0557	0.0431

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

TABLE (A3) Social Motives and Cooperation on Wikipedia: free-riders as an independent category

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
PANEL A								
Weak reciprocator	0.256	0.278	-0.0352	-0.131	-0.669	-1.317	0.129	-0.174
	(0.421)	(0.632)	(0.0608)	(0.156)	(0.566)	(0.928)	(0.0991)	(0.241)
Weak reciprocator × Superstar					1.124	2.005*	-0.262**	0.0994
					(0.806)	(1.117)	(0.124)	(0.323)
Reciprocator	0.676	0.916	-0.0773	-0.106	0.209	-0.127	0.0523	-0.197
	(0.430)	(0.641)	(0.0602)	(0.155)	(0.594)	(0.961)	(0.0963)	(0.234)
Reciprocator × Superstar					0.0272	0.623	-0.203*	0.183
					(0.833)	(1.152)	(0.121)	(0.319)
Altruist	0.788	1.371*	-0.0914	-0.217	-0.478	-0.611	0.136	-0.201
	(0.523)	(0.763)	(0.0658)	(0.186)	(0.677)	(1.119)	(0.122)	(0.301)
Altruist × Superstar					0.791	1.224	-0.338**	0.0400
					(0.917)	(1.295)	(0.144)	(0.391)
Signaler (Userpage)	6.146***	9.431***	0.00300	-0.473***	8.900***	14.39***	-0.00635	-0.288
	(0.398)	(0.604)	(0.0415)	(0.136)	(0.546)	(0.923)	(0.0843)	(0.242)
Signaler (Userpage) × Superstar					-7.205***	-12.35***	0.00943	-0.124
					(0.690)	(1.060)	(0.0948)	(0.283)
Superstar					3.399***	5.248***	0.206*	-0.255
					(0.828)	(1.122)	(0.120)	(0.314)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	647	647	477	533	647	647	477	533
adj. R^2	0.450	0.419	0.0470	0.0374	0.569	0.556	0.0543	0.0385
PANEL B								
Weak reciprocator	0.394	0.523	-0.0125	-0.0909	-0.377	-0.872	0.0557	-0.170
	(0.376)	(0.569)	(0.0532)	(0.141)	(0.483)	(0.807)	(0.0936)	(0.229)
Weak reciprocator × Superstar					0.560	1.213	-0.108	0.153
					(0.729)	(1.009)	(0.115)	(0.294)
Reciprocator	0.824**	1.174**	-0.0565	-0.0792	0.493	0.309	-0.00625	-0.177
	(0.383)	(0.573)	(0.0521)	(0.139)	(0.511)	(0.836)	(0.0906)	(0.222)
Reciprocator × Superstar					-0.485	-0.0946	-0.0768	0.189
					(0.752)	(1.034)	(0.112)	(0.289)
Altruist	1.065**	1.776***	-0.0592	-0.220	-0.144	-0.249	0.131	-0.126
	(0.465)	(0.685)	(0.0587)	(0.170)	(0.596)	(0.981)	(0.121)	(0.287)
Altruist × Superstar					0.319	0.722	-0.269*	-0.0838
					(0.834)	(1.168)	(0.138)	(0.359)
Signaler (Userpage)	6.174***	9.478***	0.00114	-0.487***	8.883***	14.39***	0.00285	-0.277
	(0.398)	(0.604)	(0.0412)	(0.136)	(0.553)	(0.934)	(0.0837)	(0.242)
Signaler (Userpage) × Superstar					-7.176***	-12.31***	-0.00760	-0.170
					(0.700)	(1.075)	(0.0943)	(0.283)
Superstar					3.906***	5.949***	0.0857	-0.242
					(0.749)	(1.003)	(0.109)	(0.286)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	647	647	477	533	647	647	477	533
adj. R^2	0.453	0.422	0.0451	0.0385	0.568	0.555	0.0485	0.0412

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). Panel A follows the very strict classification of social types defined in the main text, i.e., free-riders and altruists have mean contribution $m_i = 0$ and $m_i = 10$ across all 11 conditional contributions decisions, respectively. Panel B allows for some decision error, i.e., free-riders and altruists have mean contribution $m_i \leq 1$ and $m_i \geq 9$ across all 11 conditional contributions decisions, respectively. Control variables are the same as in the main text (not shown). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

Table (A4) Social Motives and Cooperation on Wikipedia: alternative sample of superstars

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
Reciprocator	0.454**	0.668*	-0.0346	0.00265	0.777***	0.992**	-0.0628	-0.0428
	(0.218)	(0.344)	(0.0268)	(0.0749)	(0.299)	(0.495)	(0.0435)	(0.114)
Reciprocator × Superstar					-1.020***	-1.182**	0.0522	0.0931
					(0.392)	(0.581)	(0.0546)	(0.147)
Altruist	0.528	1.056*	-0.0238	-0.0326	0.0743	0.489	0.0212	-0.0434
	(0.393)	(0.577)	(0.0429)	(0.138)	(0.443)	(0.760)	(0.0862)	(0.223)
Altruist × Superstar					-0.0570	-0.394	-0.0695	0.0694
					(0.568)	(0.834)	(0.0969)	(0.275)
Signaler (Userpage)	6.404***	9.993***	-0.0426	-0.432***	8.730***	14.08***	0.00385	-0.306
	(0.416)	(0.639)	(0.0461)	(0.146)	(0.548)	(0.917)	(0.0843)	(0.234)
Signaler (Userpage) × Superstar					-7.358***	-12.44***	-0.0808	0.0547
					(0.749)	(1.121)	(0.101)	(0.293)
Superstar					4.451***	7.131***	0.0115	-0.280*
					(0.417)	(0.588)	(0.0582)	(0.162)
Age	0.0335***	0.0421***	-0.00345***	-0.00433	0.0326***	0.0412***	-0.00337***	-0.00403
	(0.00887)	(0.0135)	(0.000887)	(0.00317)	(0.00813)	(0.0121)	(0.000881)	(0.00316)
Female	-0.959***	-1.467***	0.00983	0.0220	-1.168***	-1.820***	0.00596	0.0498
	(0.326)	(0.548)	(0.0454)	(0.125)	(0.274)	(0.462)	(0.0471)	(0.126)
Degree level	0.0580	0.147	-0.0133	0.0654**	0.0243	0.0880	-0.0143	0.0649**
	(0.0644)	(0.106)	(0.00929)	(0.0257)	(0.0590)	(0.0963)	(0.00935)	(0.0259)
Salary level	0.0371	0.0429	0.00407	0.0114	0.00365	-0.00830	0.00490	0.0138
	(0.0514)	(0.0845)	(0.00594)	(0.0178)	(0.0453)	(0.0733)	(0.00593)	(0.0180)
Risk aversion	-0.109**	-0.130*	-0.00172	-0.0127	-0.0940**	-0.102	-0.00163	-0.0129
	(0.0427)	(0.0690)	(0.00575)	(0.0170)	(0.0388)	(0.0627)	(0.00574)	(0.0170)
Nb Barnstars	0.0716***	0.0754***	-0.00317***	-0.00673	0.0656***	0.0702***	-0.00283**	-0.00238
	(0.0200)	(0.0284)	(0.00121)	(0.00490)	(0.0160)	(0.0208)	(0.00126)	(0.00452)
Size Userpage (in bytes)	-5.53e-06	-1.20e-05	1.62e-06	6.82e-06**	3.80e-06	3.63e-06	1.68e-06	6.38e-06**
	(9.80e-06)	(1.50e-05)	(1.08e-06)	(3.18e-06)	(8.25e-06)	(1.05e-05)	(1.09e-06)	(3.15e-06)
N	578	578	415	469	578	578	415	469
adj. R^2	0.482	0.446	0.0419	0.0328	0.577	0.554	0.0380	0.0371

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

TABLE (A5) Social Motives and Cooperation on Wikipedia: signaling as proportion of Barnstars displayed

	(1)	(2)	(3)	(4)
	Ln(1 + nb contrib.) *Quantity 1*	Ln(1 + bytes added) *Quantity 2*	Prop(rv w/o expl.) *Interactions*	Avg(log persist.) *Quality*
Reciprocator	-0.178	-0.109	-0.0399	0.0347
	(0.231)	(0.271)	(0.0297)	(0.0829)
Altruist	-0.143	0.0343	-0.102***	-0.193
	(0.309)	(0.325)	(0.0357)	(0.130)
Signaler (Barnstar_prop)	0.734***	0.959***	-0.0189	-0.311***
	(0.251)	(0.289)	(0.0298)	(0.0914)
Age	0.0296***	0.0267***	-0.00294***	-0.000403
	(0.00863)	(0.00865)	(0.000911)	(0.00339)
Female	-0.733**	-0.895*	0.0111	0.0102
	(0.368)	(0.533)	(0.0396)	(0.114)
Degree level	-0.0380	-0.00434	-0.0104	0.0720***
	(0.0671)	(0.0771)	(0.0100)	(0.0277)
Salary level	-0.0201	-0.00868	0.00386	0.000517
	(0.0529)	(0.0571)	(0.00618)	(0.0201)
Risk aversion	-0.00825	0.0200	0.00155	0.0104
	(0.0424)	(0.0511)	(0.00612)	(0.0167)
Nb Barnstars	0.0629***	0.0719***	-0.00286**	-0.00182
	(0.0142)	(0.0179)	(0.00117)	(0.00397)
Size Userpage (in bytes)	1.13e-05	8.62e-06	1.92e-06*	5.56e-06*
	(9.06e-06)	(1.05e-05)	(1.09e-06)	(3.32e-06)
N	308	308	292	292
adj. R^2	0.184	0.176	0.0498	0.0565

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p<0.01$, $p<0.05$ and $p<0.1$ levels, respectively.

TABLE (A6) Pairwise correlation between the average controversiality of the set of pages edited by subjects, the number of Barnstars they received, and their social signaler status

PANEL A: Casual editors				
	Ln(controversiality score)	Nb Barnstars	Signaler (Userpage)	Signaler (Barnstar)
Ln(controversiality score)	1			
Nb Barnstars	.	.		
	.			
Social signaler (Userpage)	0.366	.	1	
	($p < 0.001$)	.		
Social signaler (Barnstar)	.	.	.	.
	.	.	.	
PANEL B: Superstars				
	Ln(controversiality score)	Nb Barnstars	Signaler (Userpage)	Signaler (Barnstar)
Ln(controversiality score)	1			
Nb Barnstars	-0.019	1		
	($p = 0.719$)			
Social signaler (Userpage)	-0.054	0.291	1	
	($p = 0.318$)	($p < 0.001$)		
Social signaler (Barnstar)	0.025	0.358	0.542	1
	($p = 0.644$)	($p < 0.001$)	($p < 0.001$)	

TABLE (A7) Social Motives and Cooperation on Wikipedia: controlling for article controversiality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Avg(log persist.)
	Quantity 1	*Quantity 2*	*Interactions*	*Quality*	*Quantity 1*	*Quantity 2*	*Interactions*	*Quality*
Reciprocator	0.433**	0.595*	-0.0444*	-0.00777	0.778***	0.956**	-0.0540	-0.0417
	(0.205)	(0.315)	(0.0246)	(0.0699)	(0.294)	(0.480)	(0.0433)	(0.115)
Reciprocator × Superstar					-0.955**	-1.117**	0.0167	0.0634
					(0.378)	(0.560)	(0.0524)	(0.143)
Altruist	0.575*	0.981**	-0.0437	-0.154	0.0793	0.218	0.0932	-0.000874
	(0.333)	(0.485)	(0.0371)	(0.120)	(0.434)	(0.719)	(0.0880)	(0.219)
Altruist × Superstar					-0.0896	-0.121	-0.185*	-0.199
					(0.533)	(0.793)	(0.0945)	(0.252)
Signaler (Userpage)	5.835***	8.712***	0.00215	-0.503***	8.482***	13.26***	0.0192	-0.345
	(0.387)	(0.560)	(0.0413)	(0.134)	(0.564)	(0.900)	(0.0828)	(0.233)
Signaler (Userpage) × Superstar					-6.758***	-11.11***	-0.0233	-0.0989
					(0.722)	(1.067)	(0.0940)	(0.276)
Superstar					4.196***	6.451***	-0.000686	-0.143
					(0.376)	(0.533)	(0.0528)	(0.147)
Age	0.0316***	0.0388***	-0.00356***	-0.00311	0.0296***	0.0364***	-0.00352***	-0.00282
	(0.00837)	(0.0125)	(0.000811)	(0.00297)	(0.00752)	(0.0110)	(0.000797)	(0.00296)
Female	-0.860***	-1.203**	0.0124	0.0213	-1.018***	-1.454***	0.0106	0.0289
	(0.310)	(0.516)	(0.0423)	(0.118)	(0.260)	(0.434)	(0.0435)	(0.120)
Degree level	0.0265	0.102	-0.0120	0.0596**	0.000132	0.0573	-0.0130	0.0573**
	(0.0608)	(0.0957)	(0.00839)	(0.0241)	(0.0554)	(0.0876)	(0.00842)	(0.0242)
Salary level	-0.000617	-0.0179	0.00630	0.0126	0.00267	-0.0115	0.00715	0.0130
	(0.0492)	(0.0777)	(0.00535)	(0.0166)	(0.0425)	(0.0664)	(0.00533)	(0.0166)
Risk aversion	-0.0875**	-0.108*	-0.00178	-0.00938	-0.0689*	-0.0760	-0.00120	-0.00928
	(0.0399)	(0.0619)	(0.00528)	(0.0154)	(0.0356)	(0.0555)	(0.00525)	(0.0154)
Nb Barnstars	0.0720***	0.0802***	-0.00319***	-0.00737	0.0623***	0.0700***	-0.00266**	-0.00357
	(0.0190)	(0.0266)	(0.00115)	(0.00478)	(0.0152)	(0.0200)	(0.00117)	(0.00450)
Size Userpage (in bytes)	-3.74e-06	-8.71e-06	1.05e-06	8.87e-06***	3.26e-06	2.31e-06	1.08e-06	8.75e-06***
	(9.22e-06)	(1.38e-05)	(1.05e-06)	(3.21e-06)	(8.01e-06)	(1.07e-05)	(1.08e-06)	(3.24e-06)
Ln(edited articles controversiality score)	0.348***	0.766***	-0.0118*	0.0190	0.158**	0.460***	-0.0133*	0.0202
	(0.0819)	(0.131)	(0.00710)	(0.0238)	(0.0700)	(0.110)	(0.00734)	(0.0238)
N	647	647	477	533	647	647	477	533
adj. R^2	0.470	0.458	0.0486	0.0386	0.571	0.567	0.0527	0.0426

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

TABLE (A8) Wikipedia's Superstars: controlling for article controversiality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(1 + nb contrib.)	Ln(1 + nb contrib.)	Ln(1 + bytes added)	Ln(1 + bytes added)	Prop(rv w/o expl.)	Prop(rv w/o expl.)	Avg(log persist.)	Avg(log persist.)
	Quantity 1	*Quantity 1*	*Quantity 2*	*Quantity 2*	*Interactions*	*Interactions*	*Quality*	*Quality*
Reciprocator	-0.119	-0.144	-0.0659	-0.0994	-0.0319	-0.0313	0.0436	0.0531
	(0.220)	(0.225)	(0.261)	(0.267)	(0.0288)	(0.0290)	(0.0838)	(0.0824)
Altruist	-0.0187	-0.158	0.155	-0.0173	-0.0889**	-0.0899**	-0.130	-0.102
	(0.346)	(0.333)	(0.366)	(0.343)	(0.0380)	(0.0385)	(0.143)	(0.142)
Signaler (User page)	1.635***		2.085***		-0.0111		-0.411***	
	(0.419)		(0.511)		(0.0473)		(0.157)	
Signaler (Barnstars)		0.704***		0.947***		-0.0200		-0.307***
		(0.226)		(0.259)		(0.0280)		(0.0854)
Age	0.0281***	0.0306***	0.0247***	0.0279***	-0.00292***	-0.00293***	-0.000145	-0.000878
	(0.00812)	(0.00844)	(0.00833)	(0.00858)	(0.000921)	(0.000923)	(0.00340)	(0.00336)
Female	-0.722**	-0.820**	-0.838*	-0.960*	0.00315	0.00345	0.0169	0.0353
	(0.352)	(0.376)	(0.502)	(0.535)	(0.0413)	(0.0415)	(0.113)	(0.115)
Degree level	-0.0209	-0.0348	0.0176	-0.000196	-0.0108	-0.0107	0.0636**	0.0688**
	(0.0670)	(0.0656)	(0.0781)	(0.0756)	(0.00994)	(0.00990)	(0.0276)	(0.0277)
Salary level	-0.0348	-0.0247	-0.0293	-0.0159	0.00448	0.00414	0.00929	0.00381
	(0.0482)	(0.0508)	(0.0529)	(0.0555)	(0.00608)	(0.00606)	(0.0192)	(0.0196)
Risk aversion	0.00364	0.00819	0.0293	0.0347	0.00202	0.00215	0.00356	0.00373
	(0.0423)	(0.0427)	(0.0515)	(0.0518)	(0.00615)	(0.00615)	(0.0166)	(0.0165)
Nb Barnstars	0.0637***	0.0585***	0.0731***	0.0655***	-0.00309**	-0.00274**	-0.00366	0.000266
	(0.0145)	(0.0152)	(0.0185)	(0.0190)	(0.00120)	(0.00118)	(0.00442)	(0.00429)
Size Userpage (in bytes)	-1.89e-08	9.73e-06	-5.01e-06	7.33e-06	1.82e-06	1.78e-06*	8.11e-06**	5.99e-06*
	(9.06e-06)	(8.70e-06)	(1.14e-05)	(1.01e-05)	(1.11e-06)	(1.08e-06)	(3.71e-06)	(3.24e-06)
Ln(edited articles controversiality score)	-0.233***	-0.256***	-0.176*	-0.205**	-0.0207***	-0.0205***	0.0547*	0.0579*
	(0.0848)	(0.0826)	(0.104)	(0.102)	(0.00735)	(0.00728)	(0.0300)	(0.0305)
N	308	308	308	308	292	292	292	292
adj. R^2	0.225	0.209	0.206	0.190	0.0545	0.0560	0.0439	0.0652

The table presents OLS estimates with robust standard errors in parentheses (constant not reported). ***, ** and * denote statistical significance at the $p < 0.01$, $p < 0.05$ and $p < 0.1$ levels, respectively.

FIGURE (A3) Altruism and Cooperation: Dynamics