Computers in Human Behavior 50 (2015) 66–75

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

# Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia

Maxime Clément, Matthieu J. Guitton *

*Faculty of Medicine, Laval University, Quebec City, QC, Canada*
*Institut Universitaire en Santé Mentale de Québec, Quebec City, QC, Canada*

ARTICLE INFO

*Article history:*
Available online 11 April 2015

*Keywords:*
Community of practice
Human–machine interactions
Online collaboration
Sociotechnical systems
Virtual community
Wikipedia

ABSTRACT

With the drastic rise of social media, large-scale collaborative online projects such as Wikipedia are now dealing with incredible large amount of data. This growth forces the community to provide tremendous efforts in order to maintain the accuracy and structure of the database. To deal with such amounts of data, Wikipedia users have developed automated programs – bots – to help them to do some of the maintenance tasks. However, it is unclear how human users react to the actions of these bots. Based on a corpus of 6528 interventions (2353 different discussions) of users on talk pages of 50 bots active on English-language Wikipedia pages between January 4, 2012 and January 2, 2013, we analyzed the reactions of users depending of the characteristics of the bots' actions. Bots activity was strongly associated with the functioning Wikipedia internal community. Bots whose activity was mostly related to the work of other users (e.g. high degree of constraint or visibility) elicited more responses. By combining the different characteristics of the bots, we were able to define two opposite "ideotypes" of bots with distinct behavior: "servant bots" which mainly do repetitive and laborious work instead of human users; and "policing bots" proactively enforcing Wikipedia's guidelines and norms, which elicited more polarized responses from users (either negative or positive rather than neutral). Our results demonstrated a surprisingly high level of acceptance of bots, modulated by differential reactions in function of the actual behavior of the bots.



## 1. Introduction

The drastic increase of the size of social networks and virtual communities over the last decades has deeply affected the way humans interact online (Dunbar, 2012). Indeed, the global propensity of users for collaboration – emerging as a need to maintain efficacy – increases as the size of the virtual community increases (Yang et al., 2013). While optimizing strategies of communication within the community can help maintain social density to an extent (Guitton, 2015), ultimately, the growth of a virtual community forces its members to reduce their individual capacity of action to favor the emergence of managerial operative frameworks (Yang et al., 2013).

One of the most widely used solution that online communities have found to solve this issue was the development of specific tools which could provide assistance of human users – whether standard users or system administrators. These tools often take the form of automated programs able to operate directly in the online media, referred as "bots". Internet users have long been exposed to bots, whether they were used for entertainment in IRC, supporting the community by doing administrative work, or gathering and providing particular sets of data to users. While conversational bots are easily identified as such and are mainly used for mundane purposes, bots having explicit duties have an important role in structuring both the information available for the community, and the community itself. Indeed, their programmed scripts make them act on a systematic basis, whether the action of the bot involves material added by the bot's owner or by other users, and, more importantly, whether the human users are aware of not of the action of the bot. Therefore, due to their function and despite the fact that bots obviously do not have any form of consciousness, they strongly contribute to automate and enforce rules. In a limited community, the actions of bots would be easily identified, and the bot's owner would be considered accountable for the bot's actions. Any malfunction of the bot would immediately trigger a human reaction; the bots being then perceived as mere tools.

* Corresponding author at: Institut Universitaire en Santé Mentale de Québec, 2601 Chemin de la Canardière (F-6500), Quebec City, QC G1J 2G3, Canada. Tel.: +1 418 663 5747; fax: +1 418 663 5971.
*E-mail address:* matthieu.guitton@fmed.ulaval.ca (M.J. Guitton).

http://dx.doi.org/10.1016/j.chb.2015.03.078
0747-5632/

The situation however gets considerably more complex for larger communities. Indeed, with the considerable increase of social media and online collaborative communities and projects, the interactions between human users and bots are becoming more and more common. More than a mere increase of the magnitude of interactions, the actions of the bots over the online material increases exponentially, and in any case considerably faster than the actions of human users, due to the automated characteristics of bots' actions. Therefore, modern bots are pushing the interactions to a brand new level. In this emerging reality, the question of how human users perceive bots, and perceive the interactions with them, becomes central. More specifically, while users tend to always perceive humans as superior than bots in computer-mediated communication (Aharoni & Fridlund, 2007; Edwards, Edwards, Spence, & Shelton, 2014; Lortie & Guitton, 2011; Mowbray, 2014), the questions remains of what kind of reactions and perceptions could be expected from human users confronted with massive actions of bots, and with the need to cooperate with them more tightly.

Among the different emerging communities making a heavy use of bots, Wikipedia represents a model of choice. Wikipedia is a community-based collaborative encyclopedia, in which users can implement knowledge in a massively shared process (Fallis, 2008; Giles, 2005; Loveland & Reagle, 2013; Zlatic, Bozicevic, Stefancic, & Domazet, 2006). With the constant implementation of its content, Wikipedia is dealing with incredible large amount of data, which forces the Wikipedia community to provide tremendous efforts in order to maintain the accuracy and structure of the database (Niederer & van Dijck, 2010). To deal with such amounts of data, Wikipedia users have developed bots to help them to do some of the maintenance tasks (Geiger, 2011; Halfaker & Riedl, 2012; Müller-Birn, Herbsleb, & Dobusch, 2013; Niederer & van Dijck, 2010). With approximately 400 bots in 2010 (Niederer & van Dijck, 2010), 700 bots by September 2011, and a total of 872 bots were referenced on English Wikipedia pages on January 2, 2013, the number of bots on Wikipedia has been constantly growing. While bots were responsible for only 3–5% of the edits of the English version of Wikipedia in 2005–06, they accounted for 16.33% in 2009 (Geiger, 2009), and above 20% of the total number of editions for English Wikipedia pages in the year 2012, as compiled using Wikipedia monthly statistics. Furthermore, two thirds of new Wikipedia users receive their first message from a bot or a semi-automated algorithm (Geiger, 2014), and therefore, bots impact the retention rates of new users one way or another.

In the context of the study of human–bots interactions in virtual spaces, Wikipedia bots represent an excellent model for several reasons. First, the combination of the size of the Wikipedia community, of the sheer number of Wikipedia bots, of the size of the corpus of texts which can be edited either by humans or by bots, and of the number of actual editions performed daily on Wikipedia results in a considerable amount of potential interactions. Second, and maybe more important, is the extremely prominent social factor underlying the process of knowledge accumulation, and therefore, the implications of human–bot interactions. Indeed, the social dimension of Wikipedia is still extremely prominent, since implementation of knowledge heavily relies on community-based decision processes and "consensus" (Jahnke, 2010; Viégas, Wattenberg, & McKeon, 2007). In other words, the decision processes in Wikipedia are in fact based on social-related factors (Black, Welser, Cosley, & Degroot, 2011; Chang & Chuang, 2011). However, while clearly perceived as not equivalent in terms of social status, the Wikipedia bots nonetheless have considerably more power in terms of their action over the content of Wikipedia (power of alteration and control over the content) than a standard registered, non-administrator user (Niederer & van Dijck, 2010). Thus, the actions of bots could bypass the conventional decision process. In this context, the massively increasing actions of bots over Wikipedia content cannot be unnoticed by the community members and Wikipedia users, making Wikipedia a model of choice to study the interactions between bots and humans in an ecological large-scale setting, as well as the perception of the bots by the users.

Based on the analysis of comments on the talk pages of bots active in English Wikipedia, this study aims to characterize the interactions between users and bots. By aggregating different individual characteristics of bots, we were able to define opposite, yet complementary global profiles of bot behavior, and to demonstrate differential reactions of humans in function of these behavioral bot phenotypes. Our results lead to a better understanding of the way bots are perceived by the community, and demonstrate the importance of the reminiscence of human control over bots in their acceptation by others.

## 2. Material and methods

### 2.1. General protocol

The perception of bots by human users was evaluated on English Wikipedia active bots, via the analysis of the posts left on the bots' talk pages by users. Given that each individual language version of Wikipedia is supported by a slightly different community, the exact set of rules and conduct may slightly vary across languages. Hence, we focused on the mainstream and largest Wikipedia community, the English-language one. Bots were defined as automated software agents that perform algorithmically defined tasks involved with editing, maintenance, and administration, which can be interpreted as software tools or as managerial protocols in regards to the infrastructure (Geiger & Ribes, 2010; Müller-Birn et al., 2013; Niederer & van Dijck, 2010). The owner of a given bot is the registered Wikipedia user accountable for the actions of this bot. Each bot has a talk page, which is a background page dedicated specifically for discussions with or about the bot. In some cases, the bot's talk page is merged with the talk page of its owner. An edit consists in an edition or modification made by a user or a bot on the mainspace of Wikipedia, i.e. on Wikipedia articles. Instead, a post refers to an edit made on a talk page. The direct succession of posts concerning the same topic of interest in a talk page forms a discussion. The internal community of Wikipedia gathers the Wikipedia users who are editing with some regularity and who are at least partially aware and often part of the internal debates taking place in the collaborative project. Finally, talk page stalkers (sometimes referred as "WikiJaguar" or talk page watchers) are internal community members who survey particular talk pages (in the context of the present study, bots' talk pages) and act by answering or adding inputs to discussions on these pages, in which there were not initially directly involved.

### 2.2. Bots selection

Bots active on English Wikipedia pages between January 4, 2012 and January 2, 2013 were surveyed. Data related to bot activities were gathered from the "bots by number of edit" page of Wikipedia, which is automatically maintained by BernsteinBot, from the edit counts extracted by Toolserver (a specific software suite used by Wikipedia editors). Bot activity was assessed by calculating the difference between the number of edits done by a given bot between January 2, 2013 and January 4, 2012. A total of 170 active bots were thus identified. Pages that were directly linked to the bot talk page in order to let users post comments were retrieved, including, when relevant, the bot owner's personal talk page.

Only bots with at least 1000 edits and 12 user posts on their talk page (or the related pages) during the period of interest were

considered. Bots whose main contributions were in another language than English, or whose edits were mainly made in another language version of Wikipedia were excluded. When several bots were sharing the same talk page, only the bot with the higher edit count was considered. This resulted in a total of 51 bots. However, one of the bots (Yobot, ranking first in terms of number of edits between January 4, 2012 and January 2, 2013) was at the center of repeated conflicts, resulting in its temporary block twice over the period of sampling. Due to a problem of interpretation of bot tasks, its activity resulted in an overwhelming number of edits strongly, repeatedly, and consistently contested by the community in terms of relevance and legitimacy. Hence, the discussion pages of this bot could not be considered as a standard model of users' reactions to bots, and thus this particular bot was thus excluded from this study. Therefore, our final corpus consisted of 50 bots.

### 2.3. Bots characteristics

Bots were defined by five biphasic characteristics. The first characteristic of a bot was the nature of its owner (either an administrator of Wikipedia ("admin"), either a registered user not having administrator rights). The second characteristic was the function of the bot, defined as being mostly constraining the activity of other users (protecting from vandalism, patrolling for copyright material, warning users), or as facilitating their work (substituting templates, correcting redirect or false links). The third characteristic was the location of the modifications made by the bot, which could be either frontstage (either in article and user-space, including talk pages), or backstage (like categories, portals, help, template spaces and so on). The fourth characteristic was whether the bot acted automatically (in an opt-out way) or was only subscription-based (opt-in).

Finally, a fifth characteristic was a global evaluation of the degree of autonomy of the bot, qualified as either high (bot presenting some degree of evaluation of user contributions before acting) or low (purely factual or managerial editions, such as correcting false links or distributing email around Wikipedia users).

When bots realized multiple types of edits, characteristics were defined by the type of edits most commented by users.

### 2.4. Posts and discussions analysis

A first analysis was performed on the total number of posts on the bots associated talk pages. When discussions were not directly on the bot's talk page, but on the talk page of the bot's owner, all of the posts that were not related to the bot were excluded. If posts were not in English or were mistakenly meant for another bot, they were also excluded. Answers of bot owners and talk page stalkers were coded separately.

A second analysis was performed using only the first post of each discussion, as divided in the talk pages. Discussions were characterized as a function of the perceived intention of the initial post. Nine distinct types of users reactions were identified, reflecting either positive intention (1, acknowledgments of the bot's work; 2, help and advice requests; and 3, user informing of an action), neutral intention (4, mistake or false-positive detection; 5, enhancement suggestions; and 6, question about the bot or the edits), and negative intention (7, general criticism; 8, specific edit contestation; and 9, user justification). Examples for each of these nine types are provided in Table 1.

### 2.5. Constitution of bot ideotypes

While the analysis the five different characteristics of the bots brought key information, we noted that similar patterns of characteristics emerged on larger population of bots. Therefore, we built two opposite stereotypes, which presented specific and identifiable behavioral patterns. On the one hand, some bots mainly did repetitive and laborious work instead of human users. These bots facilitated the work of users, acted backstage, automatically (opt-out), and with a low degree of autonomy. They could be qualified as "servant bots". On the other hand, some bots proactively enforced Wikipedia's guidelines and norms. They were controlled by administrators, constraining the work of users, acting frontstage, automatically and with a high degree of autonomy. They could be referred to as "policing bots". These two types of global behavioral patterns thus represented global ideotypes of bot function and functioning.

### 2.6. Coding validity

All the initial coding was performed by a single observer, blind to the hypotheses, over a period of 40 days. Coding validity was assessed by an intra-judge evaluation: a second coding was performed on 10% of the corpus, by the same judge, with an interval period between the two coding sessions of more than 6 months.

### 2.7. Statistical analysis

Data were coded using the RQDA open source software (version 0.2-3 with R version 2.15.2), which classify the data in a SQLite database, allowing further statistical compilation from our coding. Statistical analyses were made using SPSS (version 20, IBM, Armonk), SigmaPlot (version 11, Systat Software, Chicago) and Origin (version 8.1, OriginLab, Northampton). Due to the fact that the distribution of all the recorded variables followed power laws, all data were subjected to logarithmic transformations before statistical analysis. Hence, all statistical analyses were performed on the logarithmic transformed data and not on the raw data. Analyses were performed using the non-parametric Mann–Whitney *U* test (Armitage & Berry, 1994). Simple linear regressions were performed for the nine types of users reactions in order to validate them as "positive", "neutral", or "negative" according to the highest association with the valence of the first posts of each thread. For each of the five characteristics of bots and the two ideotypes, the repartitions of the nine types of users reactions were compared using non-parametric distribution free Kolmogorov–Smirnov analyses. When applicable, results are presented as mean ± SEM.

## 3. Results

### 3.1. Characteristics of the bots population

A total of 872 bots were referenced on English Wikipedia pages on January 2, 2013. Between January 4, 2012 and January 2, 2013, 170 bots have been active, for a total of 8,985,353 edits (52,855 ± 10,437 edits per bot, median: 10,037, maximum: 975,675, minimum: 20). The distribution of their contributions follows a power law ($R^2 = 0.759$, $p < .001$, Fig. 1). A similar distribution was evidenced in our final sample of 50 bots for the total number of edits ($R^2 = 0.837$, $p < .001$ with an average of 118,360 ± 26,372 edits, median: 58,399.5, maximum: 935,971, minimum: 1828, Fig. 2A), the total number of posts on talk pages ($R^2 = 0.915$, $p < .001$ with an average of 130.56 ± 19.8 posts, median: 89.5, maximum: 668, minimum: 20, Fig. 2B), the number of posts by users on talk pages excluding the bot's owner and bot's talk page stalkers ($R^2 = 0.921$, $p < .001$ with an average of 79.52 ± 12.31 posts, median: 51.5, maximum: 433, minimum: 13, Fig. 2C), and the number of discussions on talk pages during the sampling period ($R^2 = 0.916$, $p < .001$ with an average of 47.06 ± 8.91 discussions, median: 32,

**Table 1**
Examples of users' reaction for the nine types of observed reactions.

| |
|---|
| *1. Acknowledgments of the bot's work* |
| Good catch you silly robot C6541 (T ↔ C) 21:12, 27 November 2012 (UTC) – 28 Bot |
| *2. Help and advice requests* |
| Hi, I'd need the script you used to leave a notice to all the users which participated to the SOPA's discussion, Basically I have to do almost the same for the italian one. Thank you! – Vituzzu (talk) 12:22, 17 January 2012 (UTC) – Thehelpfulbot |
| *3. User informing of an action* |
| Hi, I think that that bug in your bot is actually Pywikipedia bug (you are using this bot, right?). So, I posted the bug here: https://sourceforge.net/tracker/?func=detail&aid=3532712&group_id=93107&atid=603138 Safinaskar (talk) 10:09, 7 June 2012 (UTC) – AvicBot |
| *4. Mistake or false-positive detection* |
| It appears that RFC bot (talk · contribs) is maintaining redlinks at Wikipedia:Dashboard/Requests for comment again. (See diff) Please check to see what is causing it and if you can correct the problem. – Allen4names 15:17, 25 May 2012 (UTC) – RFC Bot |
| *5. Enhancement suggestions* |
| Regarding this diff, it would be possible to reduce the number of similar false positives by searching matched content for 'Wikipedia'.—C M B J 05:52, 14 December 2012 (UTC) – MadmanBot |
| *6. Question about the bot or the edits* |
| Should this bot revert deletion of copyvio text at Lucy Morton including addition of a copypaste template?—Preceding unsigned comment added by 82.132.232.222 (talk) 16:15, 30 July 2012 (UTC) – ClueBot NG |
| *7. General criticism* |
| Why are you spamming this nonsense to hundreds/thousands of people? Stop doing that. I didn't sign up for this crap, and I'm sure most/all others didn't either. Dream Focus 19:54, 4 September 2012 (UTC) – EdwardsBot |
| *8. Specific edit contestation* |
| Hello, there was a reason for the space. With it there Ivanhoe sorts before Ivanhoe East, which would seem to be the normal sorting convention; looks ugly the other way. Regards Crusoe8181 (talk) 07:57, 27 September 2012 (UTC) – BG19bot |
| *9. User justification* |
| I have check on this page article and in fact it is real see on this web site Official website I will add more source on this. Thank you Errorzerol (talk) 07:25, 1 September 2012 (UTC) – AnomieBot |

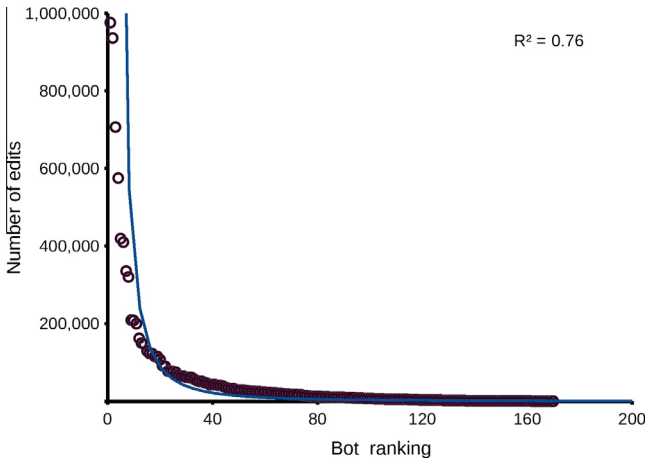


**Fig. 1.** Activity of the global bot population of the English Wikipedia. Distribution of the number of edits as a function of the bot ranking for the 170 bots active during the observation period (January 4, 2012 and January 2, 2013) on the English-language Wikipedia. This distribution follows a power law ($R^2 = 0.76$, $p < .001$).

maximum: 336, minimum: 6, Fig. 2D). At January 4, 2012, the 50 bots of our final sample were in average 1110.8 ± 92.4 days old (median: 1040, maximum: 2415, minimum: 6). At January 2, 2013, they had an average of 5.48 ± 0.95 active tasks (median: 4, maximum: 41, minimum: 1).

As mentioned earlier, some talk pages were shared between several bots. In such case, only the principal bot was counted. Therefore, 150 discussions were excluded from the 10 relevant talk pages (this phenomenon concerned 20% of the bots of our sample). Since half of the talk pages were shared between the bot and the owner of this bot (54% of our sample), and in some cases the talk page of the owner and the talk page of the bot were used concomitantly (14%), 3066 conversations were not related to the bot itself, and were thus excluded (87.6 ± 11.45 conversations per bot for the 34 bots specifically concerned by this phenomenon). Therefore, our final sampling consisted of 2353 conversations (47.06 ± 8.91 conversations per bot), for 6528 posts (130.56 ± 19.8 posts per bot and 3.14 ± 0.13 posts per conversation), including 3976 posts from general Wikipedia users (79.52 ± 12.31 posts per bot, and 1.86 ± 0.07 posts by conversation), 1939 answers from the bot's owner (38.78 ± 5.66 owner posts per bot, and 1.05 ± 0.08 owner posts per conversation), and 613 answers from talk page stalkers (12.26 ± 4.78 talk page stalker's posts per bot, and 0.18 ± 0.04 talk page stalkers' posts per conversation).

The repartitions of the bots according to the five biphasic characteristics were as follow: 26 bots were controlled by an administrator (52%) and 24 were not (48%), 17 bots constrained the users' work (34%) and 33 bots facilitated it (66%), 29 bots acted frontstage (58%) and 21 acted backstage (42%), 35 bots acted automatically (70%) and 15 were subscription-based only (30%). Finally, 16 bots had high autonomy (32%) and 34 bots had low autonomy (68%).

### 3.2. Analysis of the total number of posts

On the 3976 users' posts observed, 1346 were positive (26.92 ± 5.59 positive posts per bots, accounting for 33.7% ± 2.2% of the posts), 1704 were neutral (34.08 ± 5.21 neutral posts per bot, accounting for 42.5% ± 1.7% of the posts), and 926 were negative (18.52 ± 2.77 negative posts per bot, accounting for 23.8% ± 1.5% of the posts). The number of discussions is higher for constraining vs. facilitating (75.5 ± 22.89 vs. 32.8 ± 5.42, $p < .001$), frontstage vs. backstage (62 ± 14.64 vs. 26.4 ± 3.42, $p < .05$) and high autonomy vs. low autonomy (75.4 ± 24.39 vs. 33.7 ± 5.4, $p < .05$) (Fig. 3A). The number of total posts from users was relatively stable across characteristics, with only significantly more posts observed from constraining vs. facilitating bots (180.88 ± 45.46 vs. 104.64 ± 17.76 posts per bot, $p < .05$), and frontstage vs. backstage bots (165.24 ± 31.82 vs. 82.67 ± 11.47 posts per bot, $p < .01$). This relation is still present when considering only posts from users, and so for constraining vs. facilitating bots (113.94 ± 28.92 vs. 61.79 ± 10.34 posts per bot, $p < .01$), frontstage vs. backstage bots (100.9 ± 19.89 vs. 50 ± 6.6 posts per bot, $p < .01$) and low vs. high autonomy (64.21 ± 10.34 vs. 112.06 ± 30.69 posts per bot, $p < .05$) (Fig. 3B).

Bots controlled by administrator elicited more neutral (46.35 ± 8.97 vs. 20.79 ± 3.32 posts per bot, for bots controlled and not controlled by an administrator, respectively, $p < .05$) and

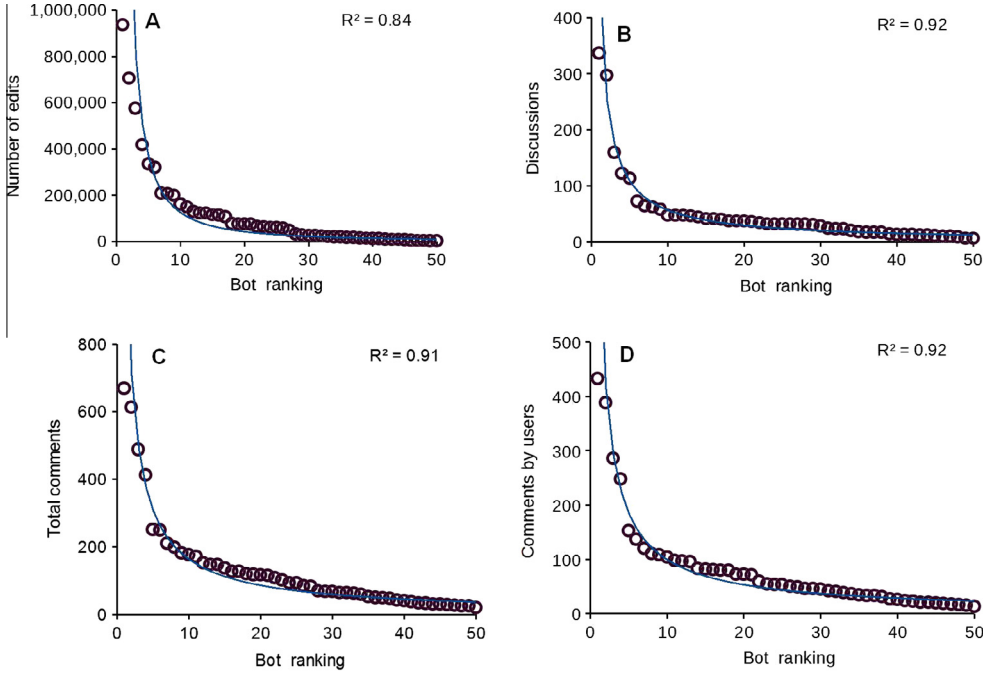


**Fig. 2.** Activity of the bot population studied. Shown are various measures of activity for the 50 bots of the studied sample. Note that all the observed distributions followed power laws. A. Distribution of the number of edits as a function of bot ranking ($R^2 = 0.84$, $p < .001$). B. Distribution of the number of discussion on bot-related talk pages as a function of bot ranking ($R^2 = 0.92$, $p < .001$). C. Distribution of the total number of interactions on bot-related talk pages as a function of bot ranking ($R^2 = 0.91$, $p < .001$). D. Distribution of the number of interactions on bot-related talk pages excluding the comments by owners and talk page stalkers, as a function of bot ranking ($R^2 = 0.92$, $p < .001$).

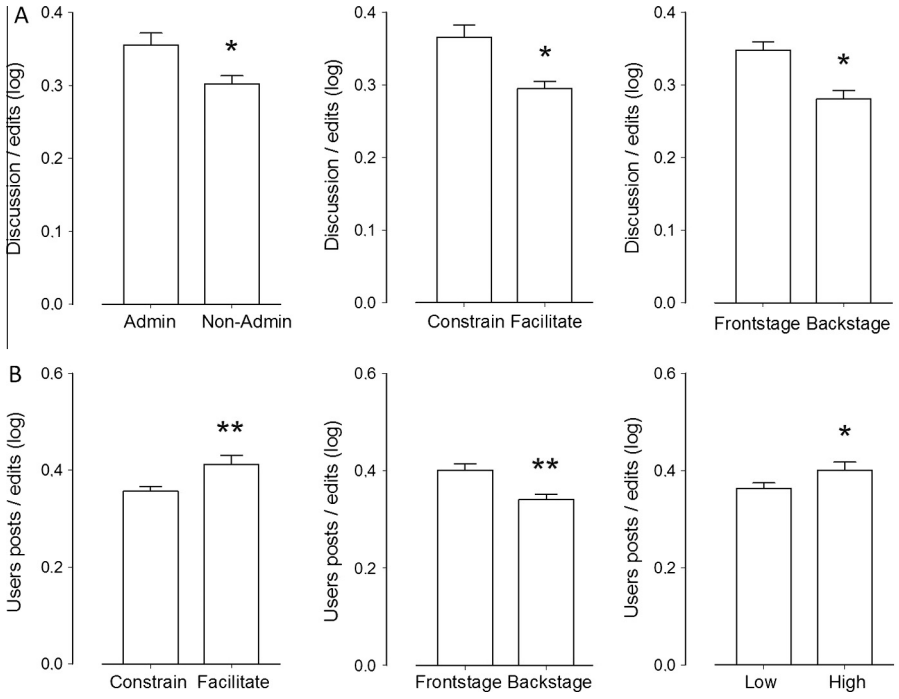


**Fig. 3.** Distributions of the interactions depending on bot characteristics. Quantification of the interactions of users on bot-related talk pages assessed by the ratio of the number of discussions/number of edits (A.) and the ratio of the number of user posts/number of edits (B.) across the five dichotomic characteristics of bots. A. Differences were observed in the ratio of the number of discussions/number of edits for three characteristics: bot's owner, bot's function, and the location of the edits of the bot. B. Differences were observed in the ratio of the number of user posts (excluding comments by owners and talk page stalkers)/number of edits for bot's function, the location of the edits of the bots, and the degree of autonomy of the bot. Statistical differences were assessed using a Mann–Whitney *U* test after a logarithmic transformation of the ratio. Differences were assessed using a Mann–Whitney *U* test. * indicates $p < .05$, and ** indicates $p < .01$.

negative posts (24.85 ± 4.71 vs. 11.67 ± 2.03 posts per bot, $p < .05$). More negative posts were observed for constraining bots than facilitating bots (28.88 ± 6.07 vs. 13.18 ± 2.39 posts per bot, $p < .01$). Bots acting frontstage elicited more polarized responses than bots acting backstage, with both more positive (35.83 ± 9.2 vs. 14.62 ± 2.29 posts per bot, $p < .01$), and negative posts

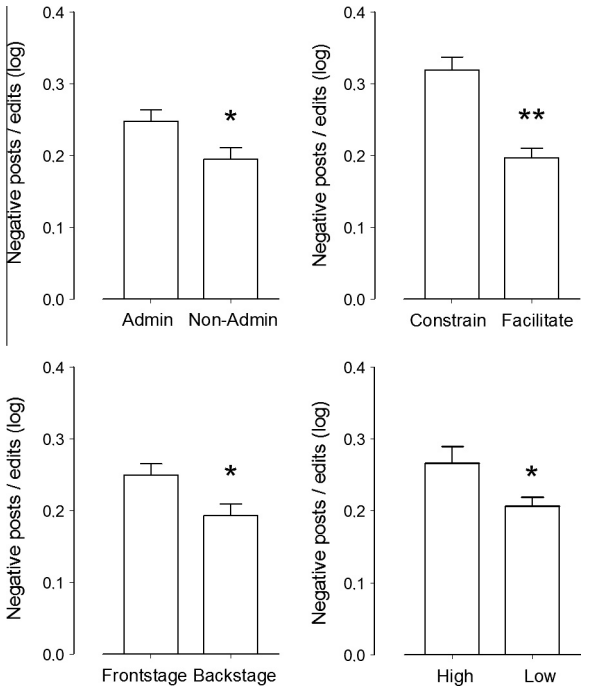


**Fig. 4.** Distribution of negative comments as a function of bots' characteristics. Significant differences were observed in the distribution of negative comments left by users on bot-related talk page for four characteristics of the bots: bot's owner (upper left), bot's function (upper right), the location of the edits (lower left), and the degree of autonomy of the bot (lower right). Statistical differences were assessed using a Mann–Whitney *U* test after a logarithmic transformation of the ratio. Differences were assessed using a Mann–Whitney *U* test. * indicates $p < .05$, and ** indicates $p < .01$.

(23.48 ± 4.4 vs. 11.67 ± 1.83 posts per bot, $p < .05$). Finally, bots with high autonomy elicited more negative posts compared to bots with low autonomy (28.81 ± 6.52 vs. 13.68 ± 2.33 posts per bot, $p < .05$) (Fig. 4). No significant difference was assessed between opt-in and opt-out bots.

Talk page stalkers were significantly more active for administrator than non-administrator (18.92 ± 8.91 vs. 5.04 ± 1.85 posts per bot, $p < .05$), frontstage bots rather than backstage bots (18.48 ± 8.03 vs. 3.67 ± 1.48 posts per bot, $p < .05$).

### 3.3. Analysis of users' types of reactions in discussions

A total of 2353 discussions was observed in our sample, including 723 positive discussions (15.26 ± 4.52 discussions per bots, 29,86% ± 2.69%), 932 neutral discussions (18.64 ± 3.06 discussions per bots, 40.78% ± 1.93%), and 657 negative discussions (13.14 ± 2.19 discussions per bots, 29.4% ± 2.15%). The average discussion contained 1.86 ± 0.47 posts, including 1.05 ± 0.54 owner posts (typically answers) and 0.18 ± 0.27 talk page stalkers interventions. The validity of the coding for the tone of the discussions was extremely high (intra-judge reliability: 95.8%).

Among the sample, 430 discussions reflected acknowledgments or praising (8.6 ± 4.07 discussions per bots, 9.7% ± 1.6%), 245 discussions were help requests (4.9 ± 0.79 discussions per bots, 11.8% ± 1.5%), 170 discussions were initiated by users informing of an action they undertook (3.4 ± 0.69 discussions per bot, 7.3% ± 0.78%), 385 discussions signalling an error (7.7 ± 1.18 discussions per bot, 21.1% ± 2.2%), 244 discussions making suggestions (4.88 ± 0.82 discussions per bot, 11.5% ± 1%), 357 questioning the logics of edits (7.14 ± 1.01 discussions per bot, 17.3% ± 1.2%), 229 were general criticism (4.58 ± 9.3 discussions per bot, 9% ± 1.1%), 198 were contestations of a specific edit (3.96 ± 0.7 discussions per bot, 9.5% ± 1.6%), and 95 user justifications (1.9 ± 0.65 discussions per bot, 2.7% ± 0.9%). Once again, the validity of the coding was extremely high (intra-judge reliability: 91.2%).

The repartitions of the types of users reactions in discussions as a function of the characteristic of the bots unveiled interesting, yet complex effects. While no overall difference in the repartition of the nine types of users reactions was observed between bots controlled by an administrator and by bots controlled by another user (Fig. 5), bots controlled by an administrator elicited significantly more responses for some types of users reactions, both positive (acknowledgment $p < .01$, user informing $p < .05$) and negative (specific contestation $p < .05$, user justification $p < .05$). In contrast, overall significant differences in the repartition of the nine types of users reactions were assessed both regarding the function of the bots ($p < .05$), and the localization of their actions ($p < .05$). Constraining bots elicited significantly more feedback than facilitating bots for one positive type of users reaction (user informing, $p < .01$), one neutral type of users reaction (questions $p < .05$), and all of the three negative types of users reactions (general criticism $p < .01$, specific contestation $p < .001$, justifications $p < .001$). Note that the frontstage modifications elicited more help requests ($p < .01$), but also questionings ($p < .05$) and general criticism ($p < .05$) than backstage modifications. Repartitions of the types of users reactions were mostly not affected by whether the bots acted automatically or were subscription-based, with users informing more ($p < .05$) and providing more justification ($p < .05$) for automatically applying bots (Fig. 5).

The pattern of reactions according to the degree of autonomy of the bot was rather similar to what was observed regarding the function the bots (Fig. 5). Overall, bots with higher degree of autonomy elicited more reactions than bots with lower autonomy ($p < .05$), with both more positive (users informing, $p < .05$) and negative reactions (general criticism $p < .05$, specific contestation $p < .05$, justifications $p < .01$).

### 3.4. Bot ideotypes

In our sample, 13 bots belonged to the "servant bot" ideotype, and 9 to the "policing bot" ideotype. Linear regressions revealed an inverse relationship for "servant bots" and "policing bots" between the number of criterions met toward the ideotype and the number of discussions observed for the bot ($R^2 = 0.28$, $p < .05$ and $R^2 = 0.25$, $p < .01$ for "servant bots" and "policing bots" respectively, Fig. 6).

"Servant bots" elicited less total interventions ($p < .01$), either for user comments ($p < .01$), for the total number of discussions ($p < .05$), for the owner responses ($p < .05$), and for talk page stalkers interventions ($p < .01$). There were less positive ($p < .01$) and negative comments ($p < .01$) than for bots not belonging to this ideotype (Fig. 7). In contrast, no significant difference was assessed on global answers for "policing bots".

Overall significant differences were found in the repartition of the nine types of users reactions between bots belonging to the ideotype and other bots, both for "servant bots" ($p < .05$), and for "policing bots" ($p < .01$). When considering the responses to the bots' actions accordingly to the nine identified types of users' reaction, "servant bots" elicited less requests for the use of the bot ($p < .05$), less edit questioning ($p < .05$) and less information on false positives or errors ($p < .05$) than bots not belonging to this ideotype (Fig. 8). "Policing bots" elicited more polarized responses, with either positive reactions (acknowledgments $p < .05$, and informing of an action $p < .05$) or negative reactions (general criticisms $p < .05$, and justification $p < .01$) significantly different than for bots not belonging to the ideotype (Fig. 8).

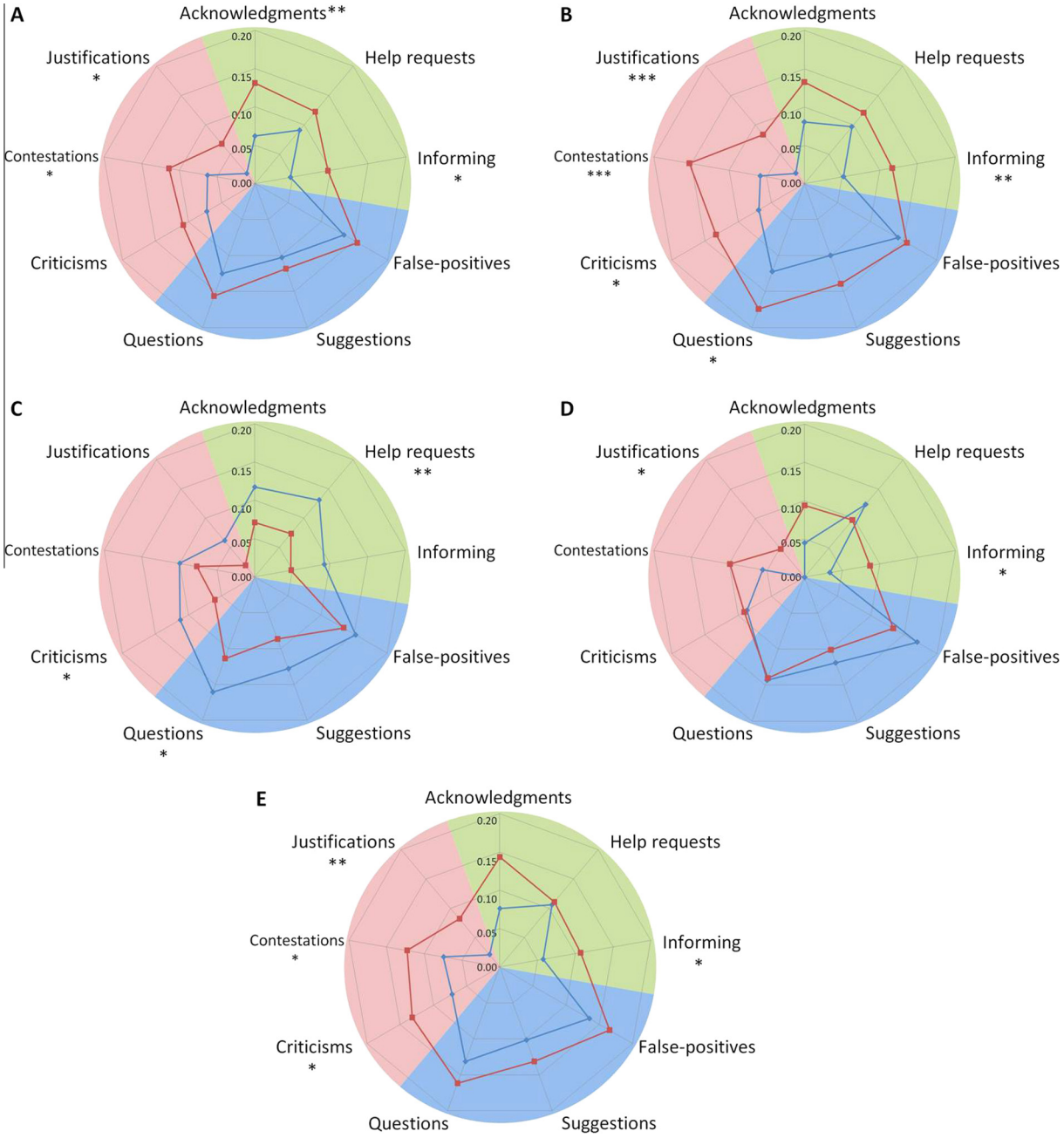


**Fig. 5.** Users reactions depending on the characteristics of the bots. Polar plots representing the proportion of the 9 identified types of user reactions to bots (3 types of positive reactions: Acknowledgments, Help requests, informing, light green; 3 types of neutral reactions: False-positives, Suggestions, Questions, light blue; and 3 types of negative reactions: Criticisms, Contestations, Justifications, light red), depending on the different characteristics of the bots. For each type of user reaction, the radial axis displays the logarithmic transformation of the ratio of the number of comments of this type of reaction/total number of edits of the bot. For each characteristic, overall differences in the repartition of types of reactions were assessed using a Kolmogorov–Smirnov test. * indicates a $p < .05$ difference and ** a $p < .01$ difference for the particular type of reaction for the characteristic (Mann–Whitney $U$ test). A. User reactions depending on the nature of the bot's owner (administrators, big dots and red line, vs. non-administrators, small dots and blue line). B. User reactions depending on the bot's function (constraining bots, big dots and red line, vs. facilitating bots, small dots and blue line, Kolmogorov–Smirnov, $p < .05$). C. User reactions depending on the location of the edits (backstage, big dots and red line, vs. frontstage, small dots and blue line, Kolmogorov–Smirnov, $p < .05$). D. User reactions depending on the degree of automatism of the bot's actions (opt-out bots, big dots and red line, vs. opt-in bots, small dots and blue line). E. User reactions depending on the degree of autonomy of the bot (high degree of autonomy, big dots and red line, vs. Low degree of autonomy, small dots and blue line, Kolmogorov–Smirnov, $p < .05$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## 4. Discussion

### 4.1. Wikipedia bots as a model of perception of human/bot interactions

While Wikipedia bots have already been the target of research interest (Geiger, 2011; Halfaker & Riedl, 2012; Müller-Birn et al., 2013; Niederer & van Dijck, 2010), the focus so far has almost always been on characterizing the putative role of the bots on structural levels – what is referred to by some authors as sociotechnical norms or system (Halfaker & Riedl, 2012; Niederer & van Dijck, 2010). The objective of the present study was fully different: we focused here on investigating how Wikipedia users reacted to the action that bots had over their work, in order to decipher the perception of the nature of the interactions between users and bots.

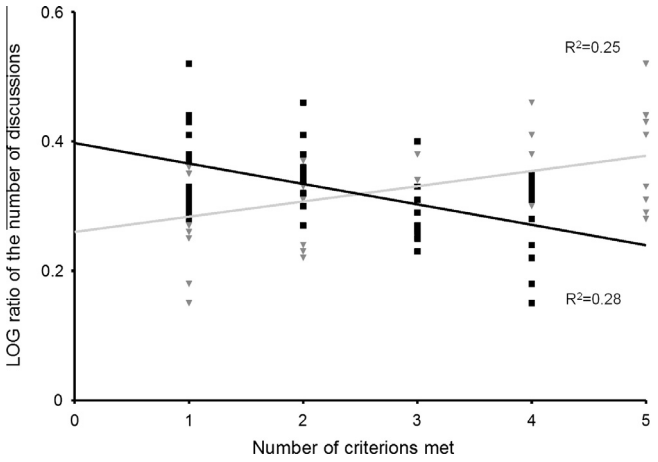


**Fig. 6.** Number of discussions as a function of the number of criterions met for each of the two identified ideotypes of bots. Linear regressions between the number of discussions (expressed as the logarithmic transformation of the ratio of the number of discussions/number of edits) and the number of criterions met for "servant" bots (black line, $R^2 = 0.28$, $p < .05$), and for "policing" bots (gray line, $R^2 = 0.25$, $p < .01$).

Our results show that not only all bots are not equal, but they also are not equally perceived. We also uncovered that bots broadly tend toward one of two stereotypes depending on the actual behavior they display, ranging from a "servant" type doing mostly repetitive minor corrections, to a more intrusive "policing" type. Bots appear to be surprisingly well accepted – even when it comes to the most intrusive ones. Furthermore, bots are largely perceived as valuable for the viability of Wikipedia, as they were consistently endorsed and defended by members of the internal community. This acceptance, which arises despite potentially ambiguous issues regarding the balance of power, might originate in the fact that the bots are still perceived as an extension of human users rather than truly independent agents.

A legitimate question would be to know whether these human–bots interactions are representative of the kind of interactions which can occur in virtual spaces, both in terms of magnitude and of quality. Investigations of Wikipedia-based large datasets have consistently demonstrated that the distribution of user-generated events followed power laws (Muchnik et al., 2013; Nazir & Takeda, 2008; Voss, 2005; Yasseri, Sumi, Rung, Kornai, & Kertész, 2012). For instance, that is the case for the number of communications between users (Muchnik et al., 2013), the number of user edits for an article (Muchnik et al., 2013), the number of edited articles by a registered user (Nazir & Takeda, 2008), the number of authors per article (Voss, 2005), the number of ingoing and outgoing links per article (Voss, 2005), the number of broken links per article (Voss, 2005), or the time between edits for a contested article (Yasseri et al., 2012). Thus, the fact that this phenomenon was also observed for our data strongly suggests that our sample reflects the actual occurrence of spontaneously generated behavior in social networks.

Despite attributing human-like qualities to bots and displaying a willingness to collaborate with them, the general perception of users about bots remains as automated servants rather than partners equal in terms of social status (Aharoni & Fridlund, 2007). Indeed, both in experimental and ecological settings, human subjects confronted to discussion bots consciously display a higher quality of inter-personal behavior when the interlocutor is identified as human vs. as a bot (Aharoni & Fridlund, 2007; Lortie & Guitton, 2011). In the present case, several elements clearly reminded users that bots were controlled by humans rather than independent agents. Indeed, more than 50% of the bots of our sample were directly controlled by Wikipedia administrators. The fact

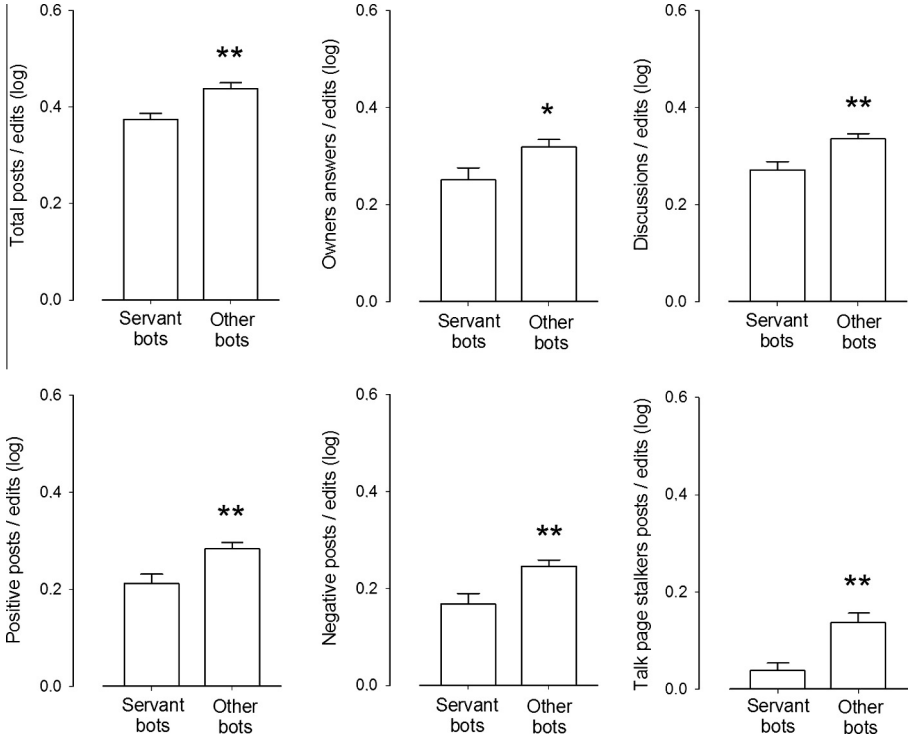


**Fig. 7.** Differences in the intensity of interactions between users and "servant" vs. other bots. A. Differences in the ratio of the total number of posts/number of edits of the bot. B. Differences in the ratio of the number of owner responses/number of edits of the bot. C. Differences in the ratio of the number of discussions/number of edits of the bot. D. Differences in the ratio of the number of positive posts/number of edits of the bot. E. Differences in the ratio of the number of negative posts/number of edits of the bot. F. Differences in the ratio of the number of talk page stalkers posts/number of edits of the bot. Statistical differences were assessed using a Mann–Whitney *U* test after a logarithmic transformation of the ratio. * indicates $p < .05$, ** indicates $p < .01$, and *** indicates $p < .001$.

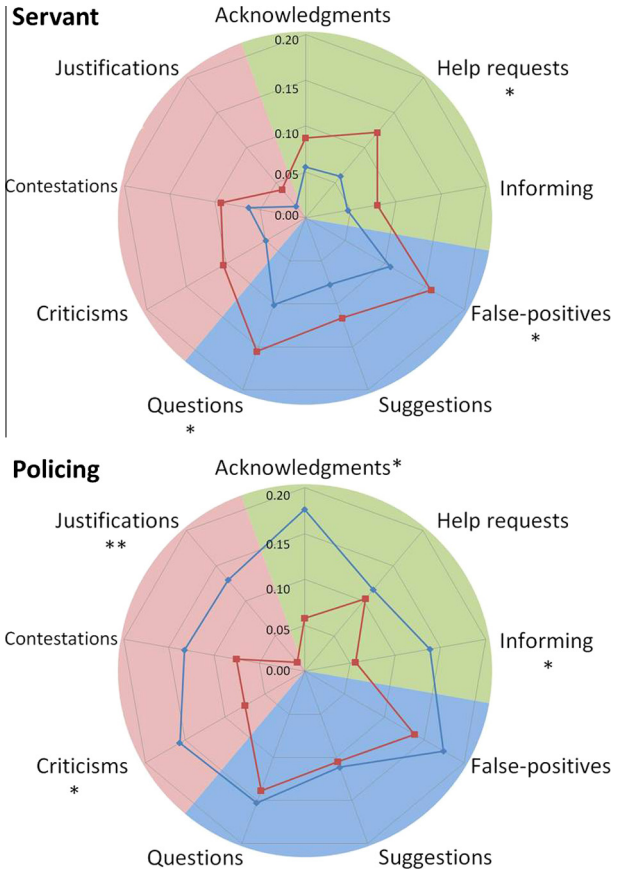


**Fig. 8.** Overall distribution of the users reactions depending on the ideotype of the bot. Polar plots representing the proportion of the 9 identified types of users reactions to bots (3 types of positive reactions: Acknowledgments, Help requests, Informing, light green; 3 types of neutral reactions: False-positives, Suggestions, Questions, light blue; and 3 types of negative reactions: Criticisms, Contestations, Justifications, light red), depending on the ideotype of the bot (servant bots, upper panel; and policing bots, lower panel). For each type of user reaction, the radial axis displays the logarithmic transformation of the ratio of the number of comments for this type of reaction/total number of edits of the bot. For each ideotype, overall differences in the repartition of types of reactions were assessed using a Kolmogorov–Smirnov test. * indicates a $p < .05$ difference and ** a $p < .01$ difference for the particular type of reaction for the characteristic (Mann–Whitney $U$ test). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that most bots discussions were at least partially redirected under the bots owner's talk page probably reinforced this perception.

### 4.2. Human users' reaction to bots actions

While human control was apparent in most of the considered bots, the possibility of massive actions of bots over the content of Wikipedia challenges this simplistic view by pushing the interactions to another level of magnitude. In such a context, would users' views of bots be limited to seeing them as purely automated servants, or would more complex perceptions occur?

Generally speaking, a very important variability was observed when considering the reactions of users to Wikipedia bots, with some bots not gathering more than 12 comments in a year. However, when considered in the light of the distributions of Wikipedia-based behavior, such variability is not surprising. The characterization of bots according to the degree of visibility of the edits they perform predicted at least partially the future interactions with the members of the community. Indeed, the bots acting frontstage concentrated most of the reactions compared to the bots acting backstage. The most negative comments from users were observed when the actions of a bot resulted in numerous very minor edits on a potentially large number of pages. Indeed, while a regular user doing such purely cosmetic edits would not be an hassle, bots doing so, with their power of treatment, can easily become overwhelming and saturate other users following a particular page with unnecessary and irrelevant information. While excluded from the present study, it is exactly what happened in the case of Yobot.

While considering bots on a continuum depending on their actualized behavior – a controlling bot vs. a helper – the users' perception evolves as a function of the role of the bots. In terms of perception, the bots closely matching the servant ideotype reflected the general stereotype of bots as automated agents. By performing repetitive and labor-intensive tasks, these bots decrease the burden born by users (Geiger, 2011; Halfaker & Riedl, 2012) and are clearly perceived as automated helpers, akin to some forms of software enhancement. Hence, these bots are the target of extremely few interactions with the community of users.

However, the situation was much more complex regarding the more invasive types of bots – the policing bots. Very often, the bots in this category systematically applied (or "force" users to apply) sets of norms which were initially supposed to be followed on a purely voluntary basis (the "ignore all rules" of the five pillars of Wikipedia). Therefore, the actions of these bots are more invasive than those of purely servant bots, and could be perceived by users as "aggressive" or, at the very least, constraining. One could have predicted that being "controlled" by a bot in terms of editorial policy would elicit massively negative reactions from human users. Indeed, as intuitively foreseen, these bots concentrated most of the interactions with users, and most of the feedbacks from the community. However, in contrast to a simplistic view, these feedbacks were not simply negative, but rather equally negative and positive.

These interactions should be read according to two opposite, yet complementary hypotheses. A first possibility would be for these bots to be perceived as external intruders by Wikipedia's contributors, thus eliciting massive criticism and rejection. Alternatively, the actions of the bots on Wikipedia pages could be perceived by users as reflecting the application of current norms, and thus, the debates over the bots' personal pages would revolves mainly around the legitimacy of the norms rather than on the actions of the bots themselves. Our data suggest that this second possibility is the correct one, since while policing bots elicited both positive and negative reactions, these reactions strongly revolve around acknowledgments or general criticisms. This might however bear a bias, since bots' owners will often use the application of the norm as an argument when facing criticisms (e.g., owners would claim that they would gladly modify the behavior of the bots once a different consensus emerges). Thus, the constant and massive implication of bot owners, administrators, and talk page stalkers that we evidenced may refrain other users from perceiving bots as too independent by constantly remaining them of their status of editorial helpers.

### 4.3. Bots and community: perception of the balance of power

The automated workforce of bots has a massive impact on the content of Wikipedia, and leads to putative questionings regarding the balance of power between bots and human users. In this context, a legitimate question is to inquire how this apparent conflict between the users' intuitive perception of bots and the reality of their power gets actualized in Wikipedia.

In Wikipedia, not all users have the same administrative powers or editorial rights. Indeed, a hierarchy can be established between the different types of users according to the actual powers they have access too, the system developers being at the top of the

pyramid while banned users are at the very bottom (Niederer & van Dijck, 2010). In this structure, the bots are just below the administrators in terms of power, but are above regular registered users, i.e., the vast majority of the active community of Wikipedia (Niederer & van Dijck, 2010). Interestingly, when a bot is controlled by a regular user instead of an administrator, the bot itself has more editorial power than its owner. It was thus not a surprise to see in the present study that administrators were highly represented among bot owners (with more than 50% of the bots of our sample directly controlled by an administrator).

With bots having major editorial power, user perceptions may be sensitive to the presence of limits to bot actions. One of the first debates on the acceptation of bots resulted in the implementation of a "no-bot" template as early as 2006, akin to a form of opt-out mechanism, allowing users to individually forbid bots to act on specific pages (Geiger, 2011; Halfaker & Riedl, 2012). Our results show that these mechanisms had the expected results. Indeed, the absence of differences between opt-in and opt-out bots in our study suggests that this parameter is not anymore a relevant criterion when interpreting the interactions between bots and users. Nonetheless, our results show that this concern about bot power is present both in the positive and the negative reactions of users, with users both informing about their actions (positive) and justifying their actions (negative) when interacting with opt-out bots. Obviously, these types of feedback decrease when interacting with opt-in bots, since they acts specifically over users demands. Finally, the fact that bots cannot revert the edition of a human having reverted the bots correction maintains the final judgement into human's hands, reducing not only the actual powers of bots, but also the potential for criticisms and negative feedbacks, as observed in our sample.

While being *a priori* minor, the numerous small, repetitive, and mostly unnoticed corrections made by the bots (particularly of the servant type) have a very high structural impact, as they concern standardizing templates or categories, on which the whole system of classification of the encyclopedia depends. Our results demonstrate that this structural role of bots is largely acknowledged by the community, as evidenced by the number of administrators and talk page stalkers coming to the rescue of bots in discussions. In this view, bots could be perceived by internal members as answering a worry of the Wikipedia community facing a massive participation of users not fully aware of the encyclopedia's standards (Butler, Joyce, & Pike, 2008).

In conclusion, our results confirm the notion that bots are currently perceived mostly as indentured collaborators and potentially valuable helpers by members of the Wikipedia community. While bots take on their own important part of the labor-intensive editorial work, they are not perceived by users as running on their own, and human control can always been felt so far. Similarly, retroactive evaluation of bot actions is the target of intense scrutiny by the community. The reactions of the community members translate a wish of having more efficient bots, more than general contestation or feelings of unease. Finally, the coherence of the reactions to bot actions could be the expression of a maturing process of the community including a reinforcement of shared sets of agreed rules.

## Acknowledgments

MJG holds a Career Grant from the "Fonds de Recherche du Québec – Santé" (FRQS). This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC – Grant Number 371644).

## References

Aharoni, E., & Fridlund, A. J. (2007). Social reactions toward people vs. computers: How mere lables shape interactions. *Computers in Human Behavior, 23*, 2175–2189. http://dx.doi.org/10.1016/j.chb.2006.02.019.

Armitage, P., & Berry, G. (1994). *Statistical methods in medical research*. Oxford: Blackwell.

Black, L. W., Welser, H. T., Cosley, D., & Degroot, J. M. (2011). Self-governance through group discussion in Wikipedia: Measuring deliberation in online groups. *Small Group Research, 42*, 595–634. http://dx.doi.org/10.1177/1046496411406137.

Butler, B., Joyce, E., & Pike, J. (2008). *Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. CHI 2008 proceedings.* Florence, Italy: ACM Press (pp. 1101–1110). Florence, Italy: ACM Press.

Chang, H. H., & Chuang, S. S. (2011). Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management, 48*, 9–18. http://dx.doi.org/10.1016/j.im.2010.11.001.

Dunbar, R. I. (2012). Social cognition on the Internet: Testing constraints on social network size. *Philosophical Transactions of the Royal Society of London: B Biological Sciences, 367*, 2192–2201. http://dx.doi.org/10.1098/rstb.2012.0121.

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior, 33*, 372–376. http://dx.doi.org/10.1016/j.chb.2013.08.013.

Fallis, D. (2008). Toward an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology, 59*, 1662–1674. http://dx.doi.org/10.1002/asi.20870.

Geiger, R. S. (2009), The social roles of bots and assisted editing programs. In *Proceedings of the 5th international symposium on wikis and open collaboration – WikiSym '09* (pp. 1–2). Orlando, Florida, USA: ACM Press. http://dx.doi.org/10.1145/1641309.1641351.

Geiger, R. S. (2014). Bots, bespoke code, and the materiality of software platforms. *Information, Communication & Society, 17*, 342–356. http://dx.doi.org/10.1080/1369118X.2013.873069.

Geiger, R. S. (2011). The lives of bots. In G. Lovink & N. Tkacz (Eds.), *Critical point of view: A Wikipedia Reader* (pp. 78–93). Amsterdam, The Netherlands: Institute of Network Cultures.

Geiger, R. S., & Ribes, D. (2010). *The work of sustaining order in Wikipedia: The banning of a vandal*. Savannah, Georgia, USA: ACM Press (pp. 117–126). Savannah, Georgia, USA: ACM Press.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*, 900–901. http://dx.doi.org/10.1038/438900a.

Guitton, M. J. (2015). Swimming with mermaids: Communication and social density in the second life merfolk community. *Computers in Human Behavior, 48*, 226–235. http://dx.doi.org/10.1016/j.chb.2015.02.004.

Halfaker, A., & Riedl, J. (2012). Bots and cyborgs: Wikipedia's immune system. *Social Computing, 12*, 79–82.

Jahnke, I. (2010). Dynamics of social roles in a knowledge management community. *Computers in Human Behavior, 26*, 533–546. http://dx.doi.org/10.1016/j.chb.2009.08.010.

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One, 6*, 1–7. http://dx.doi.org/10.1371/journal.pone.0025085.

Loveland, J., & Reagle, J. (2013). Wikipedia and encyclopedic production. *New Media & Society, 15*, 1–18. http://dx.doi.org/10.1177/1461444812470428.

Mowbray, M. (2014). Automated Twitter accounts. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 183–194). New York: Peter Lang.

Muchnik, L., Pei, S., Parra, L. C., Reis, S. D. S., Andrade, J. S., et al. (2013). Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific Reports, 3*, 1–7. http://dx.doi.org/10.1038/srep01783.

Müller-Birn, C., Herbsleb, J. D., & Dobusch, L. (2013). *Work-to-rule: The emergence of algorithmic governance in Wikipedia. C&T'13*. Munich, Germany: ACM Press (pp. 1–10). Munich, Germany: ACM Press.

Nazir, F., & Takeda, H. (2008). Extraction and analysis of tripartite relationships from Wikipedia. *Technology and Society*, 1–13.

Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society, 12*, 1368–1387. http://dx.doi.org/10.1177/1461444810365297.

Viégas, F. B., Wattenberg, M., & McKeon, M. M. (2007). The hidden order of Wikipedia. *Online Communities and Social Computing*, 445–454.

Voss, J. (2005). Measuring Wikipedia. In *Proceedings 10th international conference of the International Society for Scientometrics and Informetrics* (pp. 221–231). Stockholm, Sweden.

Yang, W., Liu, W., Viña, A., Tuanmu, M. N., He, G., et al. (2013). Nonlinear effects of group size on collective action and resource outcomes. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 10916–10921. http://dx.doi.org/10.1073/pnas.1301733110.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS One, 7*, 1–12. http://dx.doi.org/10.1371/journal.pone.0038869.

Zlatic, V., Bozicevic, M., Stefancic, H., & Domazet, M. (2006). Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E, 74*, 016115.