Title
Large-Scale Communication is More Complex and Unpredictable with Automated Bots

Journal
Journal of Communication, 70(5)

ISSN
0021-9916

Authors
Hilbert, Martin
Darmon, David

Publication Date
2020-10-01

DOI
10.1093/joc/jqaa021

Largescale Communication Is More Complex and Unpredictable with Automated Bots

Martin Hilbert* & David Darmon^

* Communication, DE Computational Social Science; University of California, Davis, CA 95616; hilbert@ucdavis.edu

^ Department of Mathematics; Monmouth University, West Long Branch, NJ 07764, USA

ABSTRACT

Automated communication bots follow deterministic local rules that either respond to programmed instructions or learned patterns. On the micro-level, their automated and reactive behavior makes certain parts of the communication dynamic more predictable. Studying communicative turns in the editing history of Wikipedia, we find that on the macro-level the overall emergent communication process becomes both more complex and less predictable. The increased presence of bots is the main explanatory variable for these seemingly contradictory tendencies. In short, individuals introduce bots to make communication more simple and predictable, but end up with a largescale dynamic that is more complex and more uncertain. We explain our results with the information processing nature of complex systems. The article also serves as a showcase for the use of information theoretic measures from dynamical systems theory to assess changes in communication dynamics provoked by algorithms.

Keywords: bots, predictability, complexity, Wikipedia, dynamical systems theory

Acknowledgements: We would like to thank Summer Mielke, who explored the adequacy of a first version of this data source as part of an undergraduate research project. We would also like to thank Ryan James, who always had an open door and helped us in countless discussions with ideas and by removing doubts.

Largescale Communication Is More Complex and Unpredictable with Automated Bots

Communication bots have become an integral part of today's online landscape. Back in the early days of the World Wide Web, Cheong (1996) still claimed to have gathered a complete list of all web-bots in some 400 pages. Shortly thereafter, their behavior was already so multifaceted that they have been declared to be a species of their own (Leonard, 1997). From 2012 to 2015, an Internet security company estimated that bots and humans had produced a quite stable 50-50 share of online traffic (Zeifman, 2015). According to a 2014 filing of Twitter, at the very least, 8.5% of all active Twitter users were bots (Costolo & Gupta, 2014), while others estimated up to 15 % (Varol et al., 2017), likely producing a multiple in terms of tweets. Bots made some 2 - 4% of all edits in the online encyclopedia Wikipedia in 2006, and more than 16% by 2010 (Geiger & Ribes, 2010). Between 2007-2013, Wikipedia bots continuously occupied 9 of the top 10 ranks of editors by edit count.[1]

There is clear evidence in the literature that communication bots have had a detectable influence on many aspects of communication processes, reaching from family and friendship relations, to democratic elections. Here we follow the tradition of "Process Theories of Communication" (Poole, 2007, p. 181; see also Berlo, 1960; Goffman, 1981; Schramm, 1955; Simmel, 1950) and ask if and how the involvement of bots affects the predictability and complexity of the unfolding communication dynamic in collective Wikipedia editing. A process-view on communication naturally lends itself to apply the tools of dynamical systems theory, so we conceptualize predictability and complexity with the time-honored summary statistics of dynamics that go back to the application of Shannon's 'mathematical theory of communication' to dynamical systems (Kolmogorov, 1958, 1959; Shannon, 1948; Sinai, 1959).

In this sense, the goal of this study is two-fold: first, we test three concrete hypotheses about changes of collective communication dynamics that occurred with increased bot involvement; second, we showcase the use of well-established measures from dynamical systems theory to characterize changes in dynamic group communication. We chose Wikipedia due to the detectable rise of bots on the platform, and the availability of large amounts of temporal data, which is needed to calculate reliable dynamical systems measures.

Theoretical Background: Algorithmification

The question of changing levels of predictability and complexity is at the core of the primary motivation for introducing bots in communication processes: automating part of the exchange. The pertinent Wikipedia article defines a Wikipedia bot as "an automated tool that carries out repetitive and mundane tasks" (Wikipedia, 2018). Automation is at the heart of computer science, as the study of computational models has been originally known as "automata studies" (Shannon & McCarthy, 1956). The word '*automaton*' is the Latinization of the Greek term for 'acting of one's own will' and refers to a self-operating agent designed to automatically execute predetermined instructions (Hopcroft & Ullman, 1979). The metaphorical 'will' of an automaton,

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

such as a bot, is known as its algorithm, which is defined as “an ordered set of unambiguous, executable steps that defines a terminating process” (Brookshear, 2009, p. 205). In short, an automaton follows a well-defined recipe to compute some result.

What automata do to human and social dynamics is what we have started to call ‘algorithmification’. The suffix ‘–fication’ comes ultimately from the Latin *facere* ‘to make, do’. So literally, it indicates to convert something (that might have been tacit before) into a quantifiable, step-by-step recipe (in the sense of Polanyi, 1966). Therefore, the algorithmification implies automating aspects of the process. Just as ‘digitalization’ introduces far-reaching consequences to the *modus operandi* of the process, reaching from network structures (Castells, 1999) to market powers (Shapiro & Varian, 1998), algorithmification leads to social dynamics that might be quite different from the previous way of doing things.

One of the main motivations for automation is to substitute procedural uncertainty with precisely defined routines. Therefore, a natural expectation of the effects of algorithmification relates to the level of predictability of the process. At the same time, communication dynamics do not happen in isolation, per definition. They are emergent phenomena, at least between two parties. Therefore, one can expect that another effect of algorithmification of communication dynamics relates to the complexity of the emergent process. We will flesh out the details of these expectations in our hypotheses, but will first build up some intuition about the concepts of predictability and complexity, which are our dependent variables.

Communication as a Dynamical System

Concepts like ‘predictability’ and ‘complexity’ are used quickly in natural language, but require a solid theoretical basis and measurable conceptualization for quantitative social science. Pursuing the goal of understanding communication dynamics, it seems natural to approach dynamical systems theory from the perspective of Shannon’s ‘mathematical theory of communication’ (1948), aka information theory (Cover & Thomas, 2006). Information theory is a branch of probability theory that conceptualizes information as the opposite of probabilistic uncertainty. Shannon famously applied it to sequences of letters in English language texts (Shannon, 1951), but expansions have quantified the characteristics of all kinds of dynamical systems (Kolmogorov, 1958, 1959; Sinai, 1959), including physical processes (Crutchfield & Feldman, 2003; Crutchfield & Young, 1989; Grassberger, 1986), biological evolution (Adami, 2012; Frank, 2012), neuroscience (Berger, 2003), and even consciousness (Seth et al., 2006; Tononi, 2011). This widely used mathematical framework comes with many advantages in the study of complex systems, maybe most importantly that it makes no assumptions as to the types of associations between variables, being able to capture multivariate nonlinear interactions as easy as simple pairwise linear correlations.

Similar dynamical systems approaches to Communication were undertaken during the 1970s. Entropy rates and Markov models were derived for phases in group decision-making (Ellis & Fisher, 1975; Krain, 1973; Poole, 1985), relational control in relationships (Fisher & Drecksel, 1983; Hawes & Foley, 1973), and talk and silence sequences in informal conversations

(Cappella, 1979). The main challenge then was computational costs and the requirement for quite extensive time series data, which was expensive to come by at the time. Today's computational power and the availability of digital trace data allows us to return to the outstanding task of characterizing communication processes as dynamics that unfold in time.

One of the insights of this framework is that the predictability and complexity of a dynamic are not necessarily and not automatically related in any predefined way. For example, in a communication between agents A and B, the following sequence of communicative turns [ABABABABABABA…] is both predictable and simple. The sequence [ABBABABBABABB…] is just as predictable (i.e. deterministic), but more memory is required to describe its main structural components (repeating AB versus ABBAB). We would say that the second sequence is more complex than the first, since its main building block contains the structural complexity of the first (i.e. AB is part of ABBAB, but not vice versa). Another way of looking at it is to say that the more complex dynamic has more variety in its sub-patterns. A two-letter parsing of both sequences identifies the structural blocks AB and BA for the first sequence, and AB, BB, and BA for the second one.

From an information theoretic perspective, complexity is, therefore, some kind of measure of the amount of identifiable patterns stored in a dynamical process (Crutchfield, 1994, 2012; Li & Vitanyi, 1997). Note that there are several dozen definitions of complexity in the literature (Lloyd, 2001), but many (if not most) are in rough agreement with this definition.

Predictions of empirical realities are rarely perfect. Following Shannon's lead, the level of predictability is then defined in terms of the remaining uncertainty that cannot be predicted from the identifiable structural components. Conditioned on the structural signatures stored in a process (its complexity), how much uncertainty remains?

In analogy to general modeling lingo, the complexity of a dynamic refers to the intricacies of its best model, while the remaining uncertainty refers to the respective error term. Complex dynamical systems exhibit all kinds of elaborate patterns situated somewhere between predictable order and unpredictable uncertainty (Crutchfield, 2012), and communication processes are no exception. What changes to complexity and predictability can be expected when algorithmifying communication dynamics with bots?

Automation to Reduce Uncertainty

Intuitively, one of the primary motivators for algorithmification is to reduce uncertainty. For one, automation increases reliability by reducing arbitrariness and increasing reliability in behavioral patterns. "Bots are predictable automatons" (Tsvetkova et al., 2017, p. 3), in short. Additionally, for modern artificial intelligence powered by machine learning, prediction itself is typically taken as the main goal (Hastie et al., 2009). Machine learning algorithms take cues from the past and anticipate the future by determining reliable patterns relating the past to the future. Algorithmification aims at exploiting both aspects: it automates specific elements of the process by replacing more aleatory aspects with reliable sub-processes, and it picks up on existing temporal signatures to anticipate future events, already incorporating a response to them into the present. Using such "prediction machines" (Agrawal et al., 2018, p. 1) in the form of an "automated tool… [for] repetitive and mundane tasks" (Wikipedia, 2018, para. 2) should then also make the overall dynamic more predictable.

Usually, bot interventions are triggered by a typical if-then clue, which results in the measurable effect on Wikipedia that "bots reciprocate much more than humans do, also at a smaller timescale" (Tsvetkova et al., 2017, p. S5). In Wikipedia, bots often provide predictable follow-up edits in response to human editing decisions. They update, redirect and repair links (e.g., after a human changed an article title), they execute newcomer greetings and add standard notification templates to explain editor decision (e.g., after a human locked a page), they update page indexes after article re-categorization, and they are involved in all kinds of rule enforcement in response to human edits, such as adding signatures, reporting potential conflict of interest, and applying property right breaches detected in one article across Wikipedia (Geiger & Halfaker, 2017; Merrill, 2015; Müller-Birn et al., 2013).

A typical example of a visible if-then reaction is an anti-vandalism bot. It reacts to the inappropriate deletion or addition of sizable amounts of text (Geiger, 2011; Geiger & Halfaker, 2013b; Geiger & Ribes, 2010). According to Geiger & Halfaker (2013), manual discoveries and reverts of vandalism take on average between one minute and 24 hours. Human assisted by bots take between five seconds and one minute. Fully automated robotic reverts almost all occur between one and five seconds. As a result of the proliferation of anti-vandalism bots, Wikipedia even published a provocative article with the self-confident title "Wikipedia: Go ahead vandalize."[2] Figure 1 visualizes how anti-vandalism bots complement human reverts.

All of these if-then actions should lead to less uncertainty, and more predictable dynamics on Wikipedia, which leads to a first research hypothesis:

H1: The involvement of bots makes the overall communication process more predictable.

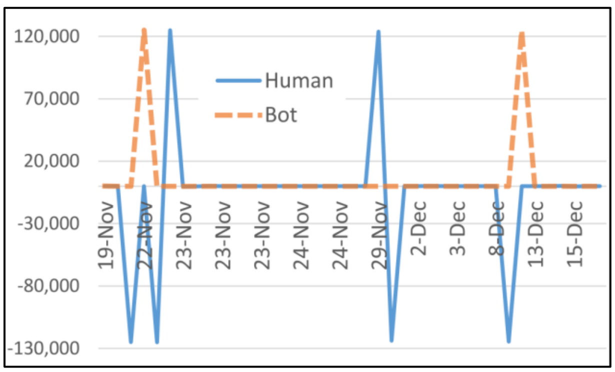


Figure 1 One month of added and deleted characters per edit in Wikipedia page 'Wikipedia'. During this month, 18 % of edits were made by bots (most edits are too small to show).

[2] https://en.wikipedia.org/wiki/Wikipedia:Go_ahead,_vandalize

Complex Bot Dynamics

The foregoing discussion of predictable if-then patterns also leaves us to expect that the variety of deterministic subcomponents of the dynamic should increase, as these can become quite complex. In the words of Tsvetkova et al. (2017, p. 3), on Wikipedia, "even relatively 'dumb' bots may give rise to complex interactions." For example, Geiger and Halfaker (2017) show that one of the most common bot activities is the redirecting of links after article titles changes, which reliably sets off a chain of bot-bot interactions. This includes bots that simplify double-redirects (which can contain dozens of steps) and other bots that update related links on all the other language versions of Wikipedia (Geiger & Halfaker, 2017). Such bot-bot interactions can represent occasional conflicts and fights among bots (Tsvetkova et al., 2017), and, in a surprising number of cases, productive and even collaborative work among bots (Geiger & Halfaker, 2017). Decades of game theory have shown that both conflictual competition and harmonious collaboration can lead to a myriad of complex patterns, which quite reliably emerge as a result of interactions among goal-oriented agents.

Our research question is formulated even more generally than bot-bot interactions, and aims at the evaluation of communication dynamics that involve both bots and humans. This results in what has been called symbiotic agency, "an agency that is implicated in the symbiotic linkages among the human and technological actors" (Neff & Nagy, 2016, p. 4927). The idea has a long and fruitful tradition in computer supported cooperative work (e.g., Orlikowski, 1992), which emphasizes that a new proxy agency emerges with the interaction of humans and bots. The new symbiotic agency does not eliminate the standalone agency of each individually. It is also usually not under the exclusive control of any of the involved parties. Qualitative research has shown that this emergent agency not only often defies the intent and expectation of the agent interacting with the bot but also of its own builder (Woolley et al., 2018). Still, being an agency, the symbiosis should lead to reliable and identifiable subcomponents of the dynamic.

Considering the effects of both bot-bot and human-bot interactions, we suggest:

H2: The involvement of bots makes the overall communication process more complex.

Direct and Indirect Bot Effects

The active Wikipedia community has seen a notable decline, from more than 52,000 in 2007 to 30,000 for 2013-2018,[3] during the same time when bots were on the rise. Several studies have hypothesized that bots' bureaucratic style of governance resulted in an ousting of humans from the platform, especially newcomers (Dobusch, 2013; Halfaker et al., 2012; Simonite, 2013). The "algorithmic tools used to reject contributions are implicated as key causes of decreased newcomer retention" (Halfaker et al., 2012, p. 664). "Being steamrollered by the newly efficient, impersonal editing machine was no fun" (Simonite, 2013, p. 5). Additionally, the automation of easier tasks traditionally used by newbies to learn the Wikipedia trade leaves only tasks for more experienced and technologically versed editors (Dobusch, 2013). It is argued that this led to a

[3] https://stats.wikimedia.org/EN/TablesWikipediaEN.htm

homogenization among Wikipedia staff, returning to the same "technical, Western, and male-dominated" demographic group as the original people who first started editing and who let loose the bots (Simonite, 2013, p. 6).

Therefore, first, any analysis of changing patterns requires being wary of potential cofounders in the form of human-machine interaction effects. It might not be the number of bots, but the constitution of human editors that leads to change in predictability and complexity. Second, the foregoing research seems to suggest a strong interaction effect between bots and editor concentration. The resulting effect on complexity and predictability is not clear. On the one hand, this interaction effects decreases human diversity (increased editor concentration), which would suggest less diverse structural patterns. On the other hand, our previous argument on symbiotic agency suggests that new diversity might emerge among experienced editors and bots. Therefore, we leave the direction of this effect as a research question to be explored, but expect:

> H3: The interaction effect of bots and editor concentration correlates with changes in predictability and complexity of the dynamic.

Method

We use a series of independent variables that characterize the editing history of articles of the English version of Wikipedia, with the average percentage of bot involvement as our focus of interest. As for dependent variables, we chose three complementary information theoretical measures, two quantifying the sequence's complexity (honoring the diversity of this concept's definition) and one for the remaining uncertainty (entropy is the axiomatically justified single variable here (Shannon, 1948)).

Data

Wikipedia provides a myriad of conversation and coordination sequences on article talk pages, interpersonal communication on user talk pages, project-wide communications for rule settings, and so much more. We chose to work with the editing sequence of articles. This is because the derivation of dynamical systems measures requires sufficiently longitudinal sequences, and making them comparable requires them to be of the same sample length (longer sequences will increase the likelihood of more structural diversity and more uncertainty by mere sampling arguments).[4] The existing data provided a good number of articles with 9,000 consecutive edits. We selected the 508th – 1,007th most revised Wikipedia articles as of October 1, 2018.[5] We end up with a good representation of similar-length articles with long editing histories[6] and 4,518,000 communicative turns from 502 articles. In order to obtain more variance between lower and

[4] The same length condition is important for statistical tests against overfitting, which we do with cross-validation.
[5] https://en.wikipedia.org/w/index.php?title=Special:MostRevisions
[6] The article of our dataset with the most revisions ('Kiss (band)') had 11,237 revisions, and the one with the least ('John F. Kennedy assassination conspiracy theories') had 9,000. We use the most recent 9,000 revisions of each, which means that in the worst case (of 'Kiss (band)'), we delete < 20% of the revisions.

higher bot levels (which is measured on average over the chosen period), we divide each editing history into two equally sized halves, leaving us with 1,004 longitudinal sequences of 4,500 edits each (we treat both halves as repeated measures in our statistical tests).

Dependent Variables and Its Measures

We decided to code editing histories in Wikipedia in terms of the number of deleted and added characters at each edit. This is the equivalent of measuring and predicting the extensiveness of a contribution during a communicative turn taken in a collective conversation among many people (measured in terms of the number of words). It is important to emphasize that the chosen dependent variable does not say anything about the quality of time invested in an edit (Geiger & Halfaker, 2013a). We also do not code for timestamps and time intervals and work only with consecutive sequences of communicative turns. We chose this measure because it is likely to capture the effects of the increased role of anti-vandalism bots on Wikipedia (Geiger, 2011; Geiger & Halfaker, 2013b; Geiger & Ribes, 2010), which is one of the main cases that informed our hypotheses (see Figure 1).

In order to use the more straightforward original measures from information theory (as compared to the more involved measures from continuous information theory (Cover & Thomas, 2006)), we convert the scalar numbers of edit sizes into categorical variables. After some sensitivity tests, we decided to use eleven bins.[7] We use two special bins to code for reverts (one bin for the initial post, and one for a subsequent reversion);[8] one bin for 0 characters (for example when someone changes upper and lower case letters); and determine the rest of the bins according to a maximum entropy principle (equal sized bins), aiming at minimizing any bias stemming from binning decisions. This results in the following bins: -[deleting >54] characters; -[54 – 6], -[6 – 0], [0], +[1 – 4], +[4 – 15], +[15 – 43], +[43 – 140], +[adding >140] characters; and two bins for a post and its subsequent revert.

We then calculate three complementary information theoretic measures on the resulting sequences of categorical data, namely predictable information E, predictive complexity C, and remaining uncertainty h, all measured in bits. Their complementary roles in dynamical systems theory is probably best explained via Figure 2a, taken from (Crutchfield et al., 2009). The Venn diagram presentation, also referred to as info-diagrams (Yeung, 1991), divides a temporal sequence into past and future. Both are quantified in terms of their entropies H (Cover &

[7] The choice of the number of bins, as well as the number of consecutive terms, is limited by the length of the available time series, since it has to be assured that each distinction among different realizations of a categorical variable provides representative statistics. Having a sequence of T = 4,500 consecutive events, and working with A = 11 bins, each group of L = 2 consecutive events should on average provide statistics from some n = 37 occurrences: $\frac{T}{n} = A^L$. This is equation gives a lower bound for expected sample size (Marton & Shields, 1994).

[8] If we identify a sum of 0 for any two post with more than 10 characters that are at most two postings apart, we mark them as reverts (one as revert addition, and one as revert deletion). We prefer this over using the revert labels provided by Wikipedia (e.g. DeDeo, 2016)), since we find them to be inconsistent. This being said, our method is susceptible to false positives. For example, it might be that n characters are posted and in the next step exactly n are deleted, even though both edits are unrelated. Checking those cases manually, we found that this would be a very rare occurrence for posts with more than 10 characters that are at most two edits apart.

Thomas, 2006). Entropy measures the uncertainty among the different events in terms of the uniformity of their distribution. One can think of obtaining statistics of the symbol sequence by using a sliding-window of a certain length across the empirically recorded event sequence and record the frequencies with which different combinations of subsequences occur. As depicted in Figure 2b, as our structural unit, we use increasing window sizes up to three consecutive communicative turns.[7] We decide on the adequate length for each individual sequence with cross-validation (one, two or three turns or letters long). The resulting subsequences are the fundamental structural units we work with for the purpose of prediction. In practice, we calculate these three measures using the Causal State Splitting Reconstruction (CSSR) algorithm for inferring ϵ-machines (epsilon-machines) (Shalizi & Klinkner, 2014) via the Python implementation from (Darmon, 2015).[9]

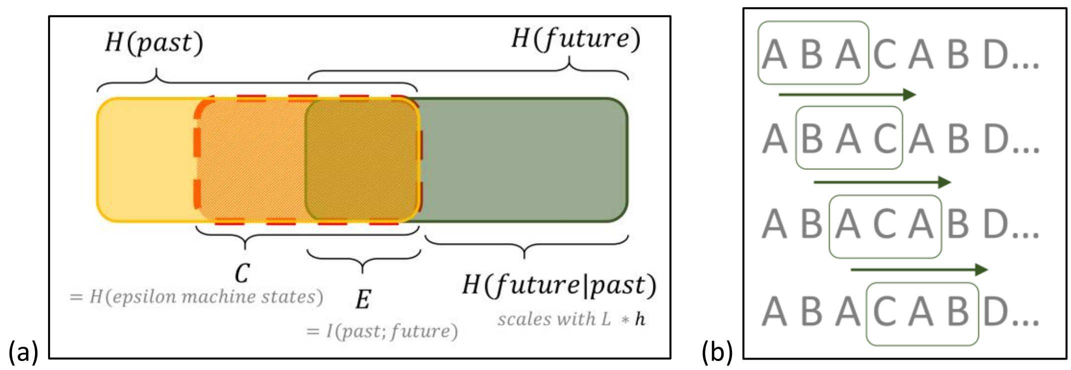


Figure 2 (a) Info-diagram relating predictive complexity (*C*), predictable information (*E*), and remaining uncertainty (*h*); (b) schematic illustration of how to obtain frequency statistics from a sequence of a categorical variable with four different bins (A, B, C, D) by employing a sliding window of length three. Obtained statistics assess the frequency of observed subsequences, the structural build blocks of the dynamic, like ‘ABA’, ‘BAC’, ‘ACA’ and ‘CAB’.

Complexity. We use the complementary measures *E* and *C* (Figure 2a).

Predictable information. The popular complexity measure we call ‘predictable information’, *E*, is also known as ‘effective measure complexity’ (Grassberger, 1986; Lindgren & Nordahl, 1988) or excess entropy in the literature (Crutchfield & Feldman, 1997), hence our symbol *E*. It is also known as ‘stored information’ (Shaw, 1984), or ‘predictive information’ (Bialek et al., 2001) (with slight differences regarding their asymptotic or time bound outlooks). It is defined as the mutual information (Cover & Thomas, 2006) between the past and the future.

[9] The main script for deriving all three measures, C, E, and h, can be found online in Python (https://github.com/ddarmon/transCSSR). A script to run it automatically over many input sequences from a csv file was added (https://github.com/3tz/transCSSR). R code for deriving E and h can be found here (https://github.com/martinhilbert/RURO-measures-and-plots).

From a perspective of information theoretic channels, it is the amount of information that the past communicates to the future (see Figure 2a). Scholars more familiar with linear statistics can see an analogy with the measure of linear autocorrelation in a time series. We call it 'predictable information' because it is the amount of information from the future that can be predicted when knowing the past.

Predictive complexity. *C* is the statistical complexity of a temporal sequence (Crutchfield & Young, 1989). *C* cannot be smaller than *E* (see Figure 2a), as it quantifies the amount of information required to predict the structure contained in *E*. In other words, *C* "captures the minimal amount of information that a process must store in order to communicate all ['predictable information', *E*]… from the past to the future" (Crutchfield et al., 2009, pp. 094101.1-.2). *C* is a practical approximation of the Kolmogorov complexity of a dynamic, as it measures the size of minimal description length of the model with maximal predictive power (Crutchfield, 2012). The model, obtained from a machine learning technique, represents the smallest size, optimally predictive, unifilar (deterministic) hidden Markov model of the dynamic (called an ϵ-machine (say 'epsilon-machine')) (Crutchfield, 1994; Shalizi & Crutchfield, 2001). According to Occam's razor and its minimum description length principle (Grunwald, 2007; Rissanen, 2010), the smallest size model has a special status in theoretical modeling. *C* quantifies the size of this statistical representation of the hidden Markov chain in terms of the entropy of its states (measured in bits). In short, the predictive complexity *C* measures the minimum amount of information required to optimally predict the future from its past. The more bits required to describe the model that makes optimal predictions, the more complex the process. *C* is zero for both deterministic and random processes, and it is maximal for stochastic processes with large memory effects.

Uncertainty. We use the entropy rate *h* to quantify how much uncertainty remains, conditioned on the identified structure.

Remaining uncertainty. Conditional entropy $H(future|past)$ measures how many bits of uncertainty are still left about the future when considering everything the past can tell us about the future (see Figure 2a). Calculated as a per symbol rate, the resulting entropy rate *h* measures the uncertainty of the next turn conditioned on the previous turns (Shannon, 1948). The overall remaining uncertainty $H(future|past)$ scales with the entropy rate with each added future symbol as $h \cdot L$ where *L* is the length of the future. An intuitive way to think about the remaining uncertainty is in terms of Fano's inequality, which provides an upper limit to the probability of error of any prediction in terms of the process' entropy rate: the larger the remaining uncertainty *h*, the larger the probability of prediction error, and the less accurate can be the predictions (Cover & Thomas, 2006; Hilbert et al., 2018).

Summing up, the predictive complexity *C* measures the amount of information required to optimally predict *E* bits of predictable information in the dynamic, while *h* quantifies the remaining uncertainty that cannot be predicted from the observed dynamic alone.

Independent Variables

Our main independent variable of interest is the level of bot involvement. It requires identifying bot accounts. The first large-scale bot operation on Wikipedia was started in 2002 with a bot that added a large number of articles about United States towns. Approximately 2,200 bot tasks were approved for use on the English Wikipedia by 2018. Wikipedia maintains a policy that requires flagging bots as such. The list had approximately 300 active bots flagged with the 'bot flag'[10] and over 400 former bots[11]. There are some known flaws with this process, most of which result in unregistered bots (Steiner, 2014). Therefore, we built our own list, based on different registries of bots that are active and deactivated, registered and unregistered, compliant and noncompliant, etc. We ended up with a list of 2,886 bots, which we verified manually if in doubt (see Supporting Information SI.3). We identified some 40 of them to be anti-vandalism bots. This small group of less than 1.5% of bots made almost 40 % of all bot edits. It hence supports our decision to code for revert activities specifically. The resulting variable '*bot-level*' measures the percentage of edits done by bots within each of our 1,004 temporal sequences of 4,500 consecutive edits (M = 3.43%, SD = 1.79%, min = 0.02%, max = 8.49%).

We need to control for several potential cofounders. First, we chose to divide the entire editing history of Wikipedia articles in half in order to both obtain more variance in terms of bot activity, and to still have long enough time sequences to derive representative dynamics. The average number of bot edits during an article's first half was 2.5%, while this increased to 4.4% for the most recent 4,500 edits of the articles recorded lifespan. However, it might also confound our analysis, as an article might be in a different stage of its lifecycle, independent from bots. We therefore control for the dummy variable '*halves*' (first half = 0; second half = 1). Given the relation between *'halves' and 'bot-level,'* we also create a variable for their interaction effect.

We furthermore control for other characteristics of the lifecycle of a Wikipedia article, including '*sum-edits*' (the length of the article in terms of the number of characters as of October 1, 2018); the categorical variable '*openness'* (distinguishing between open and protected sites, which prevents anonymous and newly registered users from editing, with open = 0 (N=702), temporarily semi-protected = 1 (N=44), indefinitely semi-protected = 2 (N=258)); and the '*growth factor*' of the second half divided by the first half (testing for the rate of growth of a page within its observed period). Finally, we measure '*editor-diversity*' (quantifying the non-uniformity of the distribution of active contributors). For this last variable, we measure the non-uniformity of the distribution of the contributions of different contributors with its discrete Shannon entropy, following common practice in biology and population ecology to measure species diversity (known as the 'Shannon-index'). Given H3, we also include the interaction effect between *'editor-diversity' and 'bot-level'* in our multivariate analysis. We end up with a total of six independent variables (IVs) and two theoretically justified interaction effects.[12] We separately test for the changing nature of each one of our three dependent variables.

[10] https://en.wikipedia.org/wiki/Special:ListUsers/bot
[11] https://en.wikipedia.org/wiki/Wikipedia:List_of_Wikipedians_by_number_of_edits/Unflagged_bots
[12] Given the inclusion of interaction effects, we center our variables '*bot-level*', '*halves'*, and '*editor-diversity*' on their respective means, reducing the risk of collinearity.

Results: Dynamical Signatures with Less and More Bots

We start with multiple regression analysis, predicting each of our information theoretic measures as dependent response variables (i.e., E, C, and h), with our eight independent predictor variables (i.e., bot-level, halves, bot-level*halves, sum-edits, growth-factor, openness, editor-diversity, bot-level*editor-diversity). We pay special attention to the role of the scalar variable bot-level. We also confirmed the identified tendencies with ANCOVAs that compare the half of the cases with lower bot-levels against the half with higher bot-levels (see Supporting Information SI.1).

Our data generally complies with several of the required assumptions for linear multivariate statistics, i.e., regarding outliers, heteroscedasticity, sample size, low collinearity (after our use of centered variables and inclusion of interaction effects), but fails the normality tests. We, therefore, bootstrap 100,000 times to create more reliable statistics (we always sample both halves of the same page together in pairs, recognizing they are not independent, but repeated measures). In general, our bootstrapped results for the estimates of the regression coefficients turn out to be more conservative than the traditional Gaussian noise estimation, so we use the bootstrap resampling for the reported significance levels at the $p < 0.01$ level.

H1. Uncertainty

A significant regression equation was found when predicting remaining uncertainty h ($F(8,995) = 10.67$, $p < 0.0005$, with an adjusted R^2 of 0.072). As shown in Figure 3a, bot-level is the strongest predictor of the model (standardized $\beta = 0.223$, $p = 0.001$). Additionally, we also find a significant interaction of bot-level with halves (std. $\beta = -0.12$, $p = 0.002$), which is negative. The negative correlation is explained by the fact that in the earlier half (the older part of the editing history), bot-level was more strongly correlated with uncertainty than in the second half (conditioned on the first half: $R^2 = 0.07$; second half: $R^2 = 0.02$). Bot-levels is a stronger predictor of the increase in uncertainty during earlier stages of an article's life cycle than during more recent periods. Editor-diversity makes a weakly significant contribution (std. $\beta = 0.138$, $p = 0.026$), which suggests that uncertainty tends to increase with more uniformity among the contributions of different editors. As such, we have to reject H1: when controlling for other effects, increasing bot-level correlates significantly with increased remaining uncertainty. According to Fano’s inequality, this lowers the limit of predictability and makes the overall process less predictable.

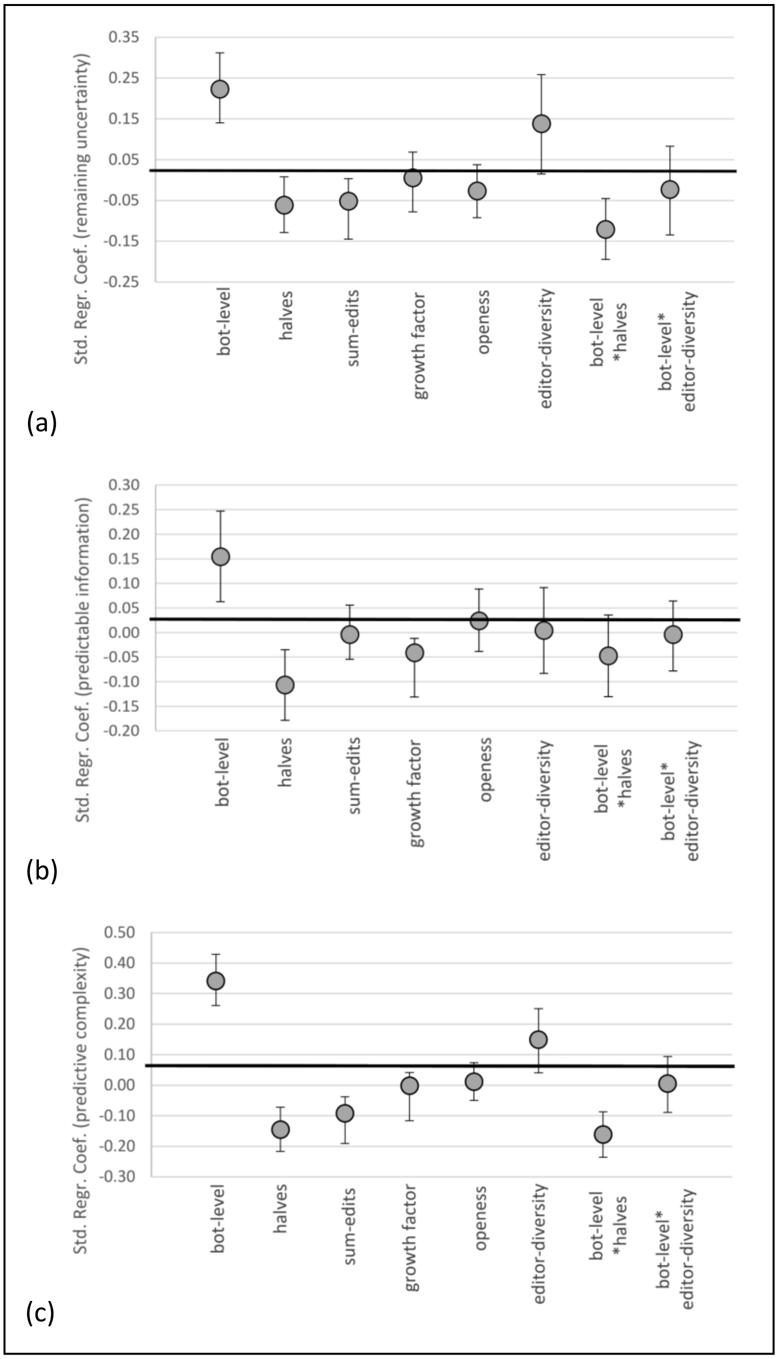


Figure 3 Standardized regression coefficient of a simultaneous regression for resampling with 100,000 bootstrapped percentile confidence intervals, with confidence bound for a marginal 95% confidence interval based on the percentile bootstrap, for (a) remaining uncertainty *h*; (b) predictable information *E*; (c) predictive complexity *C*.

H2. Complexity

Measuring complexity in terms of predictable information *E* (Figure 3b), we also find a significant regression equation (($F(8,995) = 2.5$, $p = 0.010$), with an adjusted R^2 of 0.012). Again, bot-level is our main predictor of the changing dynamic (std. $\beta = 0.155$, $p = 0.001$), followed by a negative correlation with halves (std. $\beta = -0.106$, $p = 0.004$). The growth factor of the page's content is weakly significant ($p = 0.021$). Using the predictive complexity *C* as dependent variable in a simultaneous regression with all eight IVs (Figure 3c), also results in a significant regression equation ($F(8,995) = 18.42$, $p < 0.0005$, adj. R^2 0.122). Five of our eight IVs contribute significantly to predictive complexity. The strongest predictor is again bot-level (std. $\beta = 0.341$, $p < 0.0005$), followed again by a positive relation with editor-diversity (std. $\beta = 0.149$, $p < 0.005$), and another negative relation with halves (std. $\beta = -0.145$, $p < 0.0005$). We also, once more, find another negative relation with the interaction effect bot-level*halves (std. $\beta = -0.161$, $p < 0.0005$), and a small effect of the sum-edits (std. $\beta = -0.092$, $p = 0.002$). In short, increasing bot-level is the strongest predictor of increasing both of our measures of complexity. We, therefore, cannot reject H2 that the involvement of bots makes the overall communication process more complex.

H3. Editor Concentration

First, we find that the correlation between bot-level and editor-diversity is positive, with $R^2 = 0.14$. Contrary to claims by the literature (Dobusch, 2013; Halfaker et al., 2012; Simonite, 2013), in our data, more bots went hand in hand with more editor-diversity, not less. Besides, in all three cases, the interaction effect between bot-level and editor-diversity was not significant. We find borderline stand-alone effects of editor-diversity with both increasing uncertainty *h* (Figure 3a) and predictive complexity *C* (Figure 3c). Editor diversity does seem to affect the dynamic, but not in conjunction with bot level. We, therefore, have to reject H3: we do not find support for the hypothesis that bot-influence interacts with a more or less concentrated cohort of editors.

Discussion

On Causality

It is important to point out that our tests consist of associations, which do not imply causality. Any suggested directionality in effects is derived not on an empirical, but rather on theoretical grounds. Based on this, it is likely that bots lead to changes in dynamics than the other way around: it is simply less likely that a changing dynamics in the consecutive sizes of editing dynamics were the main motivation for the introduction of bots in Wikipedia. In other words, it seems unlikely that the periods with lower bot-levels have fewer bots because the uncertainty and complexity of sequences of editing sizes did not exhibit any demand for bots, but rather because bots did not exist, or because their more limited abilities were not useful for the specific page during the period of observation. Digital algorithms are a general-purpose technology developed in many sectors across society, and studies show that the Wikipedia has adopted them because they existed in some other form to solve very particular problems (Geiger, 2011; Geiger & Ribes, 2010; Müller-Birn et al., 2013). There might be positive feedback between both, but it seems less plausible that somehow exogenously introduced changes in the sizes of editing contributions kick-started this race, but rather that exogenously provided technological possibilities started to affect editing dynamics endogenously.

This theoretical argument favors causality from bot-level to changes in dynamics: more bots lead to more uncertainty and more complexity. However, causality is extremely difficult to establish unequivocally in observational studies like ours (Pearl, 2009). The most evident limitation is the potential existence of confounding variables not included in the analysis. A much-cited study from 2012 identifies a range of changes to contribution patterns in Wikipedia related to the rise of vandalism (Halfaker et al., 2012). While we control for what we consider the most important potential confounders, we cannot assure that others might exist, such as peculiar life cycle dynamics and hidden connections among articles.

Measurement Robustness

In practice, information theoretic measures are notoriously slow to converge and, as all statistics, sensitive to different methodological decisions. We, therefore, ran a series of complementary tests using different methods to derive our measures, and using different sampling-window length, being cautious with overfitting. Besides deriving the measures from the Causal State Splitting Reconstruction (CSSR) algorithm for inferring ϵ-machines (Darmon, 2015),[9] we also calculated E and h with the Python package dit (James et al., 2018), which uses another algorithm to derive the measures. We also binned the editing sequences differently, for example, with only three bins instead of eleven bins (revert, addition, or deletion/change). Given the trade-off between representative window-length and number of categorical bins for different events,[7] we were then able to test for longer length of subsequences (following the logic outlined in Figure 2b), and run a test for a fixed word length of six consecutive edits (instead of the word length up to three consecutive edits, as before). This choice omits some information about the kinds of edits but adds information from longer temporal relations. As we show in Supporting Information SI.2, the resulting regression equations are even stronger, and the results confirm the outstanding role of increased bot-levels in the increase of uncertainty, and in predicting the increase of complexity of the editing dynamic in the form of predictable information. This gives us confidence that the general takeaway from our analysis is robust to measurement deviations.

Interpretations

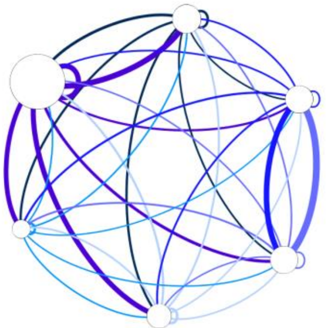
We consistently found that the level of bot involvement in Wikipedia is the best predictor to detect changing dynamics related to the sequences of the size of subsequent editing contributions. We started by theorizing that the main motivator for algorithmification consists in reducing uncertainty. Specific bots are employed to reduce uncertainty. There is no doubt that an anti-vandalism bot makes the reverts of vandalism more predictable in Wikipedia (see Figure 1). As a result, in the words of a blind reviewer of a preliminary version of our study, “if an increase in bots leads to an increase in deterministic behavior in response to the behavior of others, it seems as though under most definitions of predictability *it would necessarily be the case* that the resulting aggregate communication would be more predictable as it would have a higher proportion of content produced in a deterministic way" [emphasis added]. However, as with so many counterintuitive forms of social emergence (e.g., Schelling, 2006)), this micromotive leads to the contrary macrobehavior of the system's dynamic: largescale uncertainty is being increased.

What are some of the possible generative mechanisms that lead to more complexity and more uncertainty on the macrolevel of the communication process?

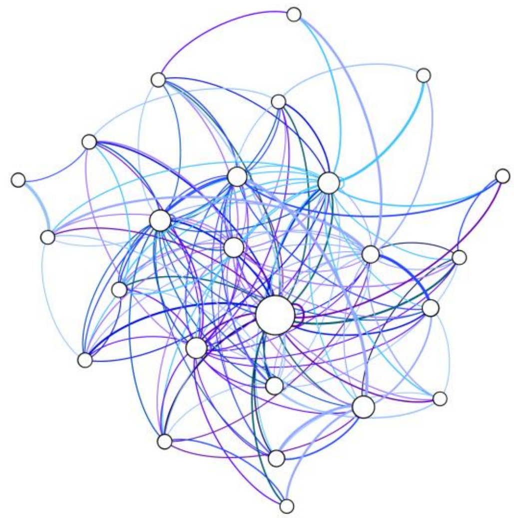
Inspired by algorithmically induced flash-crashes on the stock-market (Kirilenko et al., 2011), one candidate generative mechanism could be the emergence of more frequent extreme events that introduce volatility. Predicting the number of extreme edits (adding more than 140 or deleting more than 54 characters) with a simultaneous regression controlled by all eight IVs ($F(8,995) = 32.5$, $p < 0.0005$), it shows that bot-level is the largest predictor, but it is negatively correlated (std. $\beta = -0.404$, $p < 0.001$). The very work of anti-vandalism bots can explain this, as they aim at removing large changes. Another controlled regression shows that bot-level is also the strongest predictor for the number of reverts ($F(8,995) = 91.3$, $p < 0.0005$; std. $\beta = 0.612$, $p < 0.001$). More bots go together with more reverts and less extreme events. This should lead to more predictable overall dynamics, not to less predictable ones.

The best explanation we can offer relates to the nature of complex systems, as we usually find them in empirical reality. We derived our measures of dynamics by using a machine-learning method to fit an optimized hidden-Markov model to the temporal sequence. The minimum description length of this optimized model is our indicators for the predictive complexity C (Crutchfield & Young, 1989; Shalizi & Crutchfield, 2001). Figure 4 shows such predictive state machines for the Wikipedia entry about Mahmoud Ahmadinejad, a controversial former Iranian President.

Figure 4a shows the optimal predictive deterministic hidden Markov model for the half describing those 4,500 edits of our sample with lower bot-levels (2.3%). It has six predictive states and 66 transitions between them. Its predictive complexity is $C = 2.28$ *bits*. The accompanying entropy rate, h, is the average transition uncertainty, conditioned on the states, $h = 1.93$ *bits*. The entropy rate quantifies the average uncertainty going from state to state in the Markov chain. This remaining intrinsic uncertainty depends on the number of transitions and the uniformity or lack thereof of transition probabilities between the states. The model in Figure 4b shows the optimal model for the 4,500 consecutive edits with a higher bot level (more than twice as many, 5.5%). It has 25 predictive states and 211 transitions. Its predictive complexity is $C = 3.45$ *bits*, and its remaining uncertainty $h = 2.42$ *bits*. Complexity is higher because more states determine the model's structure. Uncertainty is higher because there are more options for the model to rattle through its structure. The more predictable structure there is, the more options there are to go about them, which leads to more uncertainty.



a)



b)

Figure 4 Visualization of predictive state machines (aka ϵ-machines) derived for https://en.wikipedia.org/wiki/Mahmoud_Ahmadinejad, modeling the (a) half with lower bot-levels (2.3%); (b) half with higher bot-levels (5.5%). Size of nodes represents the steady state probability of the state in the hidden Markov chain. Clockwise curve indicates transition directionality. Transition color codes for symbols in the original temporal sequence.

It is important to point out that there is no mathematical necessity for C and h to correlate in a specific way (Crutchfield, 2012; Shalizi & Crutchfield, 2001). There can be (and among our sample there are) dynamics with more states and larger C, but with very clearly determined transition probabilities between them and smaller h. However, these need to be designed in a very peculiar way. In empirical cases, and on average in our sample, we usually find processes that tend to follow the logic outlined here and shown in Figure 4. Among our 1,004 samples, the Person correlation coefficient between C and h is r = 0.90.

Conclusions

Individuals employ bots on the microlevel in communication processes to make aspects of the dynamic more predictable. On the macrolevel, we find that bots correlate with an unprecedented level of uncertainty, and have theoretical arguments that favor the notion that bots are the driving cause. This is counterintuitive, because, per definitions, algorithms execute predefined processes (which either have been programmed, or machine learned). This should then also lead to less uncertainty in the overall communication process, but it does not. Bot levels are the best among our candidate predictors for increased uncertainty in largescale communication dynamics in Wikipedia. In our interpretation, we argue that the main culprit for this paradox is the fact that we intuitively assume that communication processes contain a finite amount of uncertainty and that we will eventually be able to saturate it all with our predictive power. The machine learning paradigm, and its fascination with predictions, feeds many aspects of this aspiration in today's communication landscapes (Agrawal et al., 2018). However, uncertainty is not finite.

As we have seen here, algorithmification of communication processes also adds more temporal structure (predictive information E) and more computational complexity (predictive complexity C) to the communication process. Baking a more predictable structure into an ever more complex dynamic can often lead to more uncertainty, and it does in this case. What we find is less analogous to a growing body of algorithms that saturates a finite space of communicative uncertainty, but rather analogous to a growing body of algorithms that reaches further into an infinite space of uncertainty. The more new and intricate communicative patterns appear in the growing body of complex patterns, the more uncertainty is touched by the surface of this body. In other words, the more complex structure we build into communication dynamics, the more options open up for uncertainty (based on the newly built structure), and the less predictable is the process overall. As suggested here, communication bots have added a significant amount of new structural complexity to communication processes. As also suggested here, this currently leads to unprecedented amounts of uncertainty in our communication dynamics.

The implications of this finding for scholars and practitioners are manifold. As with all counterintuitive findings of social emergence, the first implication is to study it further. An important limitation for the generalization of these results is the specificity of the analyzed application (Wikipedia), and the choice of our main variables (editing size, which is mostly related to anti-vandalism bots). There is a large variety of other bot-heavy communication dynamics and many other kinds of bots on Wikipedia itself. Our presented methodology is general enough to allow future studies to verify if our results hold for other contexts.

Besides the need for empirical replications, agent-based computer simulations have turned out to be a useful methodology for illuminating emergent phenomena theoretically (Waldherr et al., forthcoming). If the finding is reconfirmed and better understood, questions arise about how to best react to increased levels of societal uncertainty. The ambition to reduce uncertainty is as old as the existence of sentient beings, who, over more than three billion years, evolved a myriad of information processes to reduce the lurking dangers of uncertain environments. The fact that the massive application of some of the most recent new ones seems to result in the opposite of their original intent leads to questions about adequate institutional responses (including governance structures and candidate mechanisms to implement rules and regulations successfully), as well as technological options (including possible meta-systems for algorithmic systems). Last but not least, it also leads to questions if and how human cognition and intelligence would be able to deal with continuous increases in complexity and uncertainty linked to the fast and omnipresent deployment of algorithmic systems.

Anonymous Online Material

An anonymous three page online appendix is available at: https://osf.io/ygk74/?view_only=6e24161edaf94bc4b9ef53869b99c201

References

Adami, C. (2012). The use of information theory in evolutionary biology. *Annals of the New York Academy of Sciences*, *1256*(1), 49–65. https://doi.org/10.1111/j.1749-6632.2011.06422.x

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Berger, T. (2003, March). Living Information Theory: The 2002 Shannon Lecture. *IEEE Information Theory Society Newsletter*, *53*(1). http://www.itsoc.org/publications/newsletters/past-newsletters/itNL0303.pdf/view

Berlo, D. (1960). *Process of Communication: An Introduction to Theory and Practice*. Harcourt School.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, Complexity and Learning. *Neural Computation*, *13*, 2001.

Brookshear, J. G. (2009). *Computer Science: An Overview* (10th ed.). Addison Wesley.

Cappella, J. N. (1979). Talk-Silence Sequences in Informal Conversations I. *Human Communication Research*, *6*(1), 3–17. https://doi.org/10.1111/j.1468-2958.1979.tb00287.x

Castells, M. (1999). *The Information Age, Volumes 1-3: Economy, Society and Culture*. Wiley-Blackwell.

Cheong, F.-C. (1996). *Internet Agents: Spiders, Wanderers, Brokers, and ’Bots* (1st Printing edition). New Riders Pub.

Costolo, R., & Gupta, M. (2014). *Form 10-Q. Twitter, Inc. Commission File Number 001-36164*. U.S. Securities and Exchange Commission (SEC). https://www.sec.gov/Archives/edgar/data/1418091/000156459014003474/twtr-10q_20140630.htm

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd Edition). Wiley-Interscience.

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, *75*(1–3), 11–54. https://doi.org/10.1016/0167-2789(94)90273-9

Crutchfield, J. P. (2012). Between order and chaos. *Nature Physics*, *8*(1), 17–24. https://doi.org/10.1038/nphys2190

Crutchfield, J. P., Ellison, C. J., & Mahoney, J. R. (2009). Time’s Barbed Arrow: Irreversibility, Crypticity, and Stored Information. *Physical Review Letters*, *103*(9), 094101. https://doi.org/10.1103/PhysRevLett.103.094101

Crutchfield, J. P., & Feldman, D. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *13*(1), 25–54.

Crutchfield, J. P., & Feldman, D. P. (1997). Statistical Complexity of Simple 1D Spin Systems. *Physical Review E*, *55*(2), R1239–R1242. https://doi.org/10.1103/PhysRevE.55.R1239

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, *63*(2), 105–108. https://doi.org/10.1103/PhysRevLett.63.105

Darmon, D. (2015). Statistical Methods for Analyzing Time Series Data Drawn from Complex Social Systems. *PhD Thesis, University of Maryland*, *Supervised by Michelle Girvan and William Rand*. https://doi.org/10.13016/M2V93N

DeDeo, S. (2016). Conflict and Computation on Wikipedia: A Finite-State Machine Analysis of Editor Interactions. *Future Internet*, *8*(3), 31. https://doi.org/10.3390/fi8030031

Dobusch, L. (2013). “Middle-aged White Guys”: Explanations for Wikipedia’s Diversity Problems. *Governance across Borders*. https://governancexborders.com/2013/08/04/middle-aged-white-guys-explanations-for-wikipedias-diversity-problems/

Ellis, D. G., & Fisher, B. A. (1975). Phases of Conflict in Small Group Development: A Markov Analysis. *Human Communication Research*, *1*(3), 195–212. https://doi.org/10.1111/j.1468-2958.1975.tb00268.x

Fisher, B. A., & Drecksel, G. L. (1983). A cyclical model of developing relationships: A study of relational control interaction. *Communication Monographs*, *50*(1), 66–78. https://doi.org/10.1080/03637758309390154

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of Evolutionary Biology*, *25*(12), 2377–2396. https://doi.org/10.1111/jeb.12010

Geiger, R. S. (2011). *The Lives of Bots* (SSRN Scholarly Paper ID 2698837). Social Science Research Network. https://papers.ssrn.com/abstract=2698837

Geiger, R. S., & Halfaker, A. (2017). Operationalizing Conflict and Cooperation between Automated Software Agents in Wikipedia: A Replication and Expansion of “Even Good Bots Fight.” *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 49:1–49:33. https://doi.org/10.1145/3134684

Geiger, R. S., & Halfaker, A. (2013a). Using edit sessions to measure participation in wikipedia. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 861–870. https://doi.org/10.1145/2441776.2441873

Geiger, R. S., & Halfaker, A. (2013b). When the Levee Breaks: Without Bots, What Happens to Wikipedia’s Quality Control Processes? *Proceedings of the 9th International Symposium on Open Collaboration*, 6:1–6:6. https://doi.org/10.1145/2491055.2491061

Geiger, R. S., & Ribes, D. (2010). The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 117–126. https://doi.org/10.1145/1718918.1718941

Goffman, E. (1981). *Forms of Talk*. University of Pennsylvania Press.

Grassberger, P. (1986). Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, *25*(9), 907–938. https://doi.org/10.1007/BF00668821

Grunwald, P. D. (2007). *The Minimum Description Length Principle*. The MIT Press.

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*. https://doi.org/10.1177/0002764212469365

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2 edition). Springer.

Hawes, L. C., & Foley, J. M. (1973). A Markov analysis of interview communication. *Speech Monographs*, *40*(3), 208–219. https://doi.org/10.1080/03637757309375798

Hilbert, M., James, R. G., Gil-Lopez, T., Jiang, K., & Zhou, Y. (2018). The Complementary Importance of Static Structure and Temporal Dynamics in Teamwork Communication. *Human Communication Research*, *44*(4), 427–448. https://doi.org/10.1093/hcr/hqy008

Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to Automata Theory, Languages and Computation* (1st edition). Addison-Wesley Publishing Company.

James, R. G., Ellison, C. J., & Crutchfield, J. P. (2018). dit: A Python package for discrete information theory. *Journal of Open Source Software*, *3*(25), 738.

Kirilenko, A. A., Kyle, A. S., Samadi, M., & Tuzun, T. (2011). The Flash Crash: The Impact of High Frequency Trading on an Electronic Market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1686004

Kolmogorov, A. N. (1958). A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR.*, *119*(861–864), 2.

Kolmogorov, A. N. (1959). Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR.*, *124*(4), 754–755.

Krain, M. (1973). Communication as a Process of Dyadic Organization and Development. *Journal of Communication*, *23*(4), 392–408. https://doi.org/10.1111/j.1460-2466.1973.tb00957.x

Leonard, A. (1997). *Bots: The Origin of New Species* (1st edition). Hardwired.

Li, M., & Vitanyi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer.

Lindgren, K., & Nordahl, M. G. (1988). Complexity measures and cellular automata. *Complex Systems*, *2*, 409–440.

Lloyd, S. (2001). Measures of complexity: A nonexhaustive list. *IEEE Control Systems*, *21*(4), 7–8. https://doi.org/10.1109/MCS.2001.939938

Marton, K., & Shields, P. C. (1994). Entropy and the Consistent Estimation of Joint Distributions. *The Annals of Probability*, *22*(2), 960–977. https://doi.org/10.1214/aop/1176988736

Merrill, B. (2015). The Bots Who Edit Wikipedia (And The Humans Who Made Them). *MakeUseOf*. https://www.makeuseof.com/tag/bots-edit-wikipedia-humans-made/

Müller-Birn, C., Dobusch, L., & Herbsleb, J. D. (2013). Work-to-rule: The Emergence of Algorithmic Governance in Wikipedia. *Proceedings of the 6th International Conference on Communities and Technologies*, 80–89. https://doi.org/10.1145/2482991.2482999

Neff, G., & Nagy, P. (2016). Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, *10*(0), 17.

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*(3), 398–427. https://doi.org/10.1287/orsc.3.3.398

Pearl, J. (2009). *Causality*. Cambridge University Press.

Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.

Poole, M. S. (1985). Tasks and interaction sequences: A theory of coherence in group decision-making interaction. In R. L. Street & J. N. Cappella (Eds.), *Sequence and Pattern in Communicative Behaviour* (pp. 206–224). Edward Arnold.

Poole, M. S. (2007). Generalization in Process Theories of Communication. *Communication Methods and Measures*, *1*(3), 181–190. https://doi.org/10.1080/19312450701434979

Rissanen, J. (2010). *Information and Complexity in Statistical Modeling* (Softcover reprint of hardcover 1st ed. 2007). Springer.

Schelling, T. C. (2006). *Micromotives and Macrobehavior*. W. W. Norton & Company.

Schramm, W. L. (1955). *The process and effects of mass communication*. University of Illinois press.

Seth, A. K., Izhikevich, E., Reeke, G. N., & Edelman, G. M. (2006). Theories and measures of consciousness: An extended framework. *Proceedings of the National Academy of Sciences*, *103*(28), 10799–10804. https://doi.org/10.1073/pnas.0604347103

Shalizi, C. R., & Crutchfield, J. P. (2001). Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics*, *104*(3–4), 817–879. https://doi.org/10.1023/A:1010388907793

Shalizi, C. R., & Klinkner, K. L. (2014). Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences. *ArXiv:1408.2025 [Cs, Stat]*. http://arxiv.org/abs/1408.2025

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423, 623–656. https://doi.org/10.1145/584091.584093

Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Tech. Journal*, *30*, 50–64.

Shannon, C. E., & McCarthy, J. (Eds.). (1956). *Automata Studies.* Princeton University Press.

Shapiro, C., & Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy* (1st ed.). Harvard Business Press.

Shaw, R. (1984). *The Dripping Faucet as a Model Chaotic System*. Aerial Press.

Simmel, G. (1950). *The Sociology of Georg Simmel* (K. H. Wolff, Trans.). Simon and Schuster.

Simonite, T. (2013). The Fight to Save Wikipedia from Itself. *MIT Technology Review*, *Intelligent Machines*. https://www.technologyreview.com/s/520446/the-decline-of-wikipedia/

Sinai, Y. G. (1959). On the notion of entropy of dynamical systems. *Doklady Akademii Nauk*, *124*(4), 768–771.

Steiner, T. (2014). *Bots vs. Wikipedians, Anons vs. Logged-Ins*. https://doi.org/10.1145/2567948.2576948

Tononi, G. (2011). The Integrated Information Theory of Consciousness: An Updated Account. *Archives Italiennes de Biologie*, *150*(2/3), 56–90. https://doi.org/10.4449/aib.v149i5.1388

Tsvetkova, M., García-Gavilanes, R., Floridi, L., & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. *PLOS ONE*, *12*(2), e0171774. https://doi.org/10.1371/journal.pone.0171774

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017, May 3). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Eleventh International AAAI Conference on Web and Social Media*. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587

Waldherr, A., González-Bailón, S., & Hilbert, M. (forthcoming). Agent-Based Modeling for Communication Research. *Communication Methods and Measures*, *Special Issue*. https://think.taylorandfrancis.com/cfp-ah-hcms-si-agent-based-modeling/

Wikipedia. (2018). *Wikipedia:Bots*. https://en.wikipedia.org/wiki/Wikipedia:Bots#Bot_policy,

Woolley, S. C., Shoery, S., & Howard, P. (2018). The Bot Proxy. In Z. Papacharissi (Ed.), *A Networked Self and Platforms, Stories, Connections* (pp. 59–76). Routledge.

Yeung, R. W. (1991). A new outlook on Shannon's information measures. *IEEE Transactions on Information Theory*, *37*(3), 466–474. https://doi.org/10.1109/18.79902

Zeifman, I. (2015). *2015 Bot Traffic Report: Humans Take Back the Web, Bad Bots Not...* Imperva Incapsula. https://www.incapsula.com/blog/bot-traffic-report-2015.html

Supporting Information

Largescale Communication Is More Complex and Unpredictable with Automated Bots

Martin Hilbert* & David Darmon^

* Communication, Computational Social Science; University of California, Davis, CA 95616, USA; hilbert@ucdavis.edu

^ Department of Mathematics; Monmouth University, West Long Branch, NJ 07764, USA; ddarmon@monmouth.edu

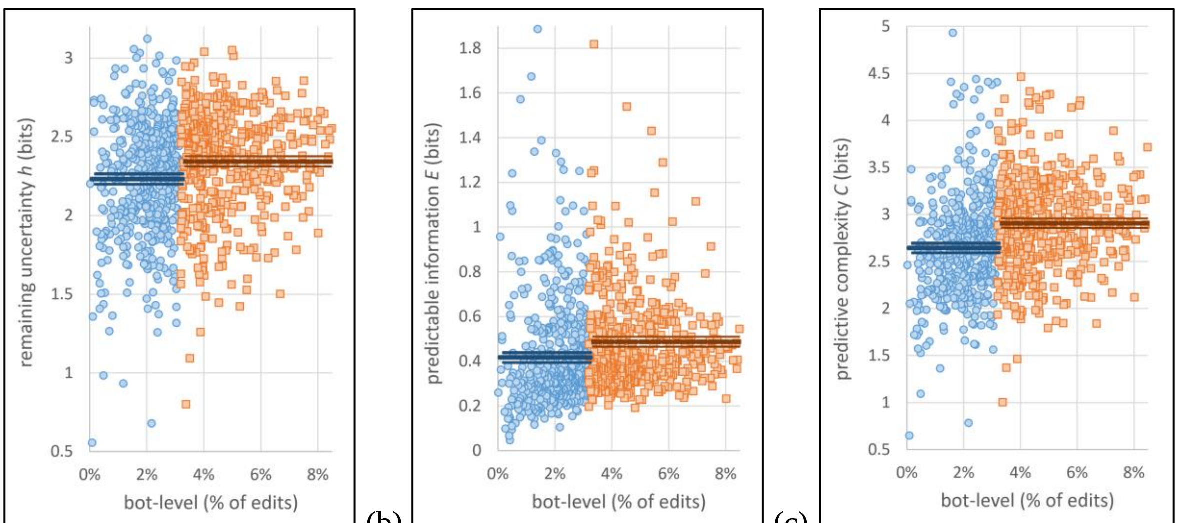
This online appendix is also available at:
https://osf.io/ygk74/?view_only=6e24161edaf94bc4b9ef53869b99c201

Contents

SI.1 Visual Reconfirmation and ANCOVA

In addition to the multiple regression results presented in the main article, we also ran different ANCOVAs, which naturally leads to similar results, given the close relation between both methods. The analysis loses some information, since the scalar variable of bot-level is dichotomized into a binary dummy of halves with lower (0) and higher bot-levels (1) as the fixed factor, but it has the benefit that it leads to visualizations that illustrate the orders of magnitude and the variance involved in the observed tendencies (see Figure SI.1). In general, we find 2.0% of bot involvement for the half of the data with lower levels, and more than twice as much, 4.9%, for the half with higher bot-levels. For all three measures, bot-level continued to be the strongest contributor, for h with ($F(1,995) = 18.81$, $p < 1.0E\text{-}04$, bootstrapped estimated marginal mean $M_{lowbot} = 2.232$; $M_{highbot} = 2.343$), for E with ($F(1,995) = 16.72$, $p < 1.0E\text{-}04$, bootstr. mean $M_{lowbot} = 0.417$; $M_{highbot} = 0.487$), and for C with ($F(1,995) = 48.77$, $p < 1.0E\text{-}04$, bootstr. mean $M_{lowbot} = 2.644$; $M_{highbot} = 2.907$). Figure 4 shows that the bootstrapped 95% confidence intervals on the estimated marginal mean values are not overlapping in any of the cases.

Figure SI.1: ANCOVA distinguishing among lower and higher bot levels; estimated marginal means with 95% confidence intervals from 1,100 bootstrap samples, of (a) remaining uncertainty *h*; (b) predictable information *E*; (c) predictive complexity *C*.

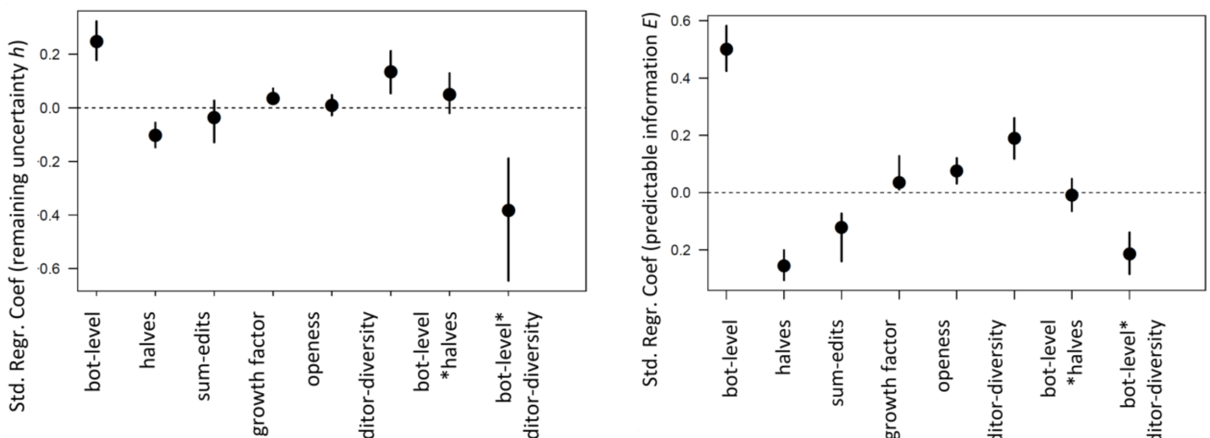


(a) (b) (c)

SI.2 Robustness of Measurements

For our tests in the main article, we derived our information theoretic dependent variables from the Causal State Splitting Reconstruction (CSSR) algorithm for inferring ϵ-machines (Shalizi & Klinkner, 2014) via the Python implementation from (Darmon, 2015). Here we calculated predictive information *E* and remaining uncertainty *h* with the Python package dit (James, Ellison, & Crutchfield, 2018), which uses another algorithm to derive the measures. Also, instead of binning the underlying editing size of our temporal sequences in eleven bins, here we binned them using only three bins (revert, addition, or deletion/change). Given the smaller alphabet, we then had enough statistical power to test for longer window lengths of subsequences, which define the basic structural components of our dynamics (aka, the recurrent words, made of the alphabet of events).[13] After testing for several window lengths, in the following, we used a test for a fixed word length of six consecutive edits (instead of the word length up to three, as we did for the larger alphabet). Using less bins omits some information about the kinds of edits, but using longer windows adds information from longer temporal relations.

Figure SI.2: Standardized regression coefficient of a simultaneous regression for resampling with 100,000 bootstrapped percentile confidence intervals, with confidence bound for a marginal 95% confidence interval based on the percentile bootstrap, for (a) remaining uncertainty *h*; (b) predictable information *E*; both based on ternary alphabet and fixed window length of six, calculated with (James et al., 2018).



a) b)

The resulting multiple regressions are stronger than the ones presented in the main article, with ($F(8,995) = 100.0$, $p < 1.0E-04$, with an adjusted R2 of 0.4412) for *h*, and ($F(8,995) = 109.5$, $p < 1.0E-04$, with an adjusted R2 of 0.4638). As shown in Figure SI.2, for both cases, bot-levels are the single best predictor variable of the changing dynamics. It is notable that the interaction between bot-level and editor-diversity is significantly negative in both cases, while the individual contribution of both variables is positive. The negative correlation of the interaction effect is explained by the fact that in the half with lower bot levels, editor-diversity was much stronger correlated with the dynamical measures than in second half: for *h*, conditioned on the half with lower bot level: $R^2 = 0.37$; half with higher bot level: $R^2 = 0.002$; and for *E*, conditioned on the half with lower bot level: $R^2 = 0.36$; half with higher bot level: $R^2 = 0.009$). With less bots, the

[13] The choice of the number of bins, as well as the number of consecutive terms, is limited by the length of the available time series, since it has to be assured that each distinction among different realizations of a categorical variable provides representative statistics. Having a sequence of T = 4,500 consecutive events, and working with A = 11 bins, each group of L = 2 consecutive events should on average provide statistics from some n = 37 occurrences: $\frac{T}{n} = A^L$. This is equation gives a lower bound for expected sample size (Marton & Shields, 1994).

effect of more uniformity among editors and their contributions has a much larger effect on increased uncertainty and complexity.

Besides, in this version, we find that most variables have a significant effect. Notably editor-diversity has a positive effect on the dynamic (more uniformity among contributors and the number of their contributions, more uncertainty and complexity), and halves has a negative effect (more recent edits have less uncertainty and complexity, independent from bot-level).

We ran other tests, with word length of five, etc, and grew increasingly confident with our results, independent from measurement differences. We decided to present the version in the main article, because it is one of the more conservative versions, and considers an adequate level of detail (e.g. distinction among 11 bins).

SI.3 List of bots

From main article: In 2018, there were approximately 2,200 bot tasks approved for use on the English Wikipedia. Wikipedia maintains a policy that requires flagging bots as such. The list had approximately 300 active bots flagged with the ‘bot flag’ and over 400 former bots . There are some known flaws with this process, most of which result in unregistered bots (Steiner, 2014). Therefore, we built our own list, based on different registries of bots that are active and deactivated, registered and unregistered, compliant and noncompliant, etc. We ended up with a list of 2,886 bots, which we verified manually if in doubt.

Here goes the list:

"Quote"bot; (1.VSNCT)Bot; .anacondabot; .snoopybot.; ^demonBot2; ____ robot ; _______ (bot); __-bot; _bot; 0–9; 10.4.1.125; 115 Glitch; 123Hedgebot456; 126Bot; 1VeertjeBot; 1zzz; 28bot; 3RRBot; 718 Bot; 718 Bot ; 762bot; 7SeriesBOT; A bot called Bob; A bot called Bob ; A4bot; AABot; AAlertBot; AarghBot; Ab.awbot; Abbot; AbiBot; Abisys.bot; AbiyoyoBot; AboHeidiBot; ABot; Abotzi; Absconditus bot; ACBot; AccReqBot; AccReqBot2; Acebot; AdambroBot; Adas bot; Addbot; Addihockey10; Addihockey10 (automated); Addihockey10 (automated) ; AddihockeyBot; AdertBot; Adlerbot; AdminStatsBot; Admrbot; AdQ Bot; Adrián Neves Bot; AEBot; AFCbot; AFCHBuddy; AFD Bot; AFD Bot ; AfDBot; AfdlBot; AfDStatBot; Afkbot; Africanus; AftabBot; AGbot; Agtbot; AHbot; Ahechtbot; Aibot; AIDbot; AilurophobiaBot; AiraBot; AiuwBot; Akasenbot; AkBot; AkeronBot; AkhtaBot; Aksibot; Aktionsbot; Al Maghi Bot; Alaibot; AlaibotToo; AlanBOT; Albambot; ALBot; Albotim; Alch Bot; AldnonymousBOT; Ale jrb bot; Ale jrb bot ; Ale jrb bot/welcomebot archive; Alecs.bot; AlekseyBot; Aleph Bot; Alertbot; AlessioBot; Aleth Bot; AlexandriaBot; Alexbot; AlexNewArtBot; Alfiobot; AliBot; Alirezabot; AlkhemaBot; Allanbot; AllaSmackBot; AlleborgoBot; AllieBot; Almabot; AlnoktaBOT; AlohaBOT; AloysiusLiliusBot; Alph Bot; Alph Bot ; Alphachimpbot; AlphamaBot; AlphamaBot3; AlptaBot; AlptaSandboxBot; A-lú-mih-bot; AlyBot; AMABot; AMahoney bot; Amalthea; Amalthea (archivebot); Amalthea (bot 2); Amalthea (bot); AmaraBot; AMbot; Ameenbot; AmeliorationBot; AmerigoBOT; Amical-bot; AmigaBot; Amirobot; Amolbot; AmphBot; amSpadeBot; An13saBot; Analphabot; Anchor Link Bot; Anchor Link Bot ; AnchorBot; AndersBot; Andrea105Bot; AndreasJSbot; Andrebot; Andrewbot; Andrewbot~frwiki; Andriy.vBot; Android Mouse Bot; Android Mouse Bot ; Android Mouse Bot 2; Android Mouse Bot 2 ; Android Mouse Bot 3; Android Mouse Bot 4; AndrzeiBOT; AndymwBot; AngBot; AngelBot; Angria77Bot; Anibot; AnkitAWB; AnkitBot; AnnabelsBot; AnomieBOT; AnomieBOT II; AnomieBOT III; AnomieBOT/UserpageMain; Anonymoususerdaily; AntiAbuseBot; AntidereBot; Antigng-bot; Antime Bot I ; Antime Bot I~enwiki; Antischmitzbot; AntiSpamBot; AntiVandalBot; AntonyB-Bot; Anybot; AP.BOT; APersonBot; Aplikasi-Bot; Apocatequil; Apocatequil~enwiki; Apocatequil~frwiki; Apostrobot; APPERbot; Aptenodibot; Aptenodybot; Aptenodybot A; AraBot; ArbComBot; Arbitrarily0Bot; ArcheoBot; Archi-bot; Archivbot; Archive 'o' matic; ArkBot; ArkyBot; ArmadilloProcessBot; Armanbot; ArmbrustBot; ArnauBot; ARSBot; ArshavaBot; ArthurBot; ArticleAlertbot; ArticleAlertingBot; ArticleGrinder; ArticleListBot; ArticlesForCreationBot; Artsiom91Bot; ARVBot; AryanBot; ArystanbekBot; ASammourBot; AsenineBot; AsgardBot; AstaBOTh15; AStarBot; AstraleBot; AstRoBot; AsturiBot; AsuraBot; AswnBot; Atarubot; AT-bot; Athikhun.suwBOT; Atobot; AttoBot; AttributionBot; AudeBot; AushulzBot; AusTerrapinBotEdits; Autobot; AutocracyBot; Autorit_Bot; Autotitlechange; AVBOT; AVdiscuBOT; Averaterbot; AvicBot; AvicBot2; AvocatoBot; AWeenieBot; AwOcBot; AzaBot; AZatBot; AZatBot~enwiki; AZPR; B.Zsoltbot; B0ttle; BaBoT; Babylon5; Babylon5~enwiki; BacDiveBot; BackfireBot; BackfireBotII; BadCitationBot; BAGBot; Baifito; BaldBot; Balmung0731-AWB; BanjoBot; BanjoBot 2.0; BankuBot; BannerBot; BArchBot; BarkBot; BartBotje; BaseBot; BasketBallStatsBot; BatreeqbotPro; BattyBot; BazookaBot; B-bot; BD2412bot; Bean49Bot; Beastie Bot; Beastie Bot ; Beastie Bot/Stage One; Beau.bot; BécherBot; Beckhamwong04; BegbertsBot; Begemot-Bot; BekoBot; BekusBot; Bellezzasolo Bot; BendelacBOT; Bender the Bot; BeneBot2008; BenjaBot; BenjBot; BenjiBot; BenoniBot; BenoniBot~enwiki; BentongBot; BenzolBot; BepBot; BeriBot; BernsteinBot; Bersibot; BetaBot;

BetacommandBot; BetBot; BetBot~enwiki; Beussonbot; BezkingBot; BezkingBot-Link; BG19bot; Bgbot; BHGbot; Bibcode Bot; Bigsus-bot; Bikabot; Bill william compton (Bot); BimBot; BinBot; Biobot; BioPhotoBot; BirthdayBot; Bitobot; BJBot; BjornNbot; BlackBot; BlackDart; BlackDemon2; BlackpoolFCBot; BlakesBot; BlevintronBot; Blockbot; Blood Oath Bot; Blood Oath Bot ; BloodFire; Bloodsource/bot; BLPWatchBot; Bluebot; BlueRobot; BlueRobot~enwiki; Bobergerbot; BoBot; Bocianski.bot; BodhisattvaBot; BogBot; BokimBot; BolatbekBot; Bolling26; Bollins Cerrname/BollinsBot; Bonabot; Bonobot; Bookbot; BorgardeBot; Borgxbot; BoricuaBot; bot; Bot FH; Bot FH ; Bot for Justice; Bot for Justice ; Bot noteszzz; Bot que revierte; Bot Tox; Bot van der Hoorn; Bot van der Hoorn ; Bot0612; Bot1058; Bot11598; Bot24; Bot523; Bota47; Botalex; BotAniste; BotanyBot; BOTarate; Botarel; Bot-chan~enwiki; Bot-chan-II; BotChristophe; BotCompuGeek; Botcrux; BotdeSki; BotDHojalata; Boten Anna; BOTepho; BoTieno; BOTijo; BOTirithel; Bot-Jagwar; BotKung; Botlaf; BotMultichill; BotMultichillT; BotMyShinyMetalAss; BotNautilus; BotNinja; Botodo; Botones; BoToulousain; BoToX; Botpankonin; BOTpolicia; BotPuppet; Botrie; Botryoidal; Bot-Schafter; BotSottile; BotStats; BOT-Superzerocool; BotteHarry; Bottuzzu; BOT-Twm Crys; Botx; Botz~enwiki; BoulaurBot; BoxCrawler; BracketBot; BrackiBot; BraincricketBot; BrainistBot; BrandonBot; BreakBot; Brest-bot; Broadbot; BrokenAnchorBot; Brother Abbott; Brother Abbott ; BrownBot; BryanBot; BsherrAWBBOT; BSqure; BU RoBOT; Bublebot; Bub's wikibot; BücherBot; Bulwersator: bot; BumlaBot; ButkoBot; BWBot; Byrialbot; BŽcherBot; C1PB8; C-3POrao; CaBot; CactusBot; CAGBot; CahyoBot; Caiserbot; CakeBot; CalakBot; CamelBot; CanaryBot; CanisRufus; CapacciBot; CapitalBot; Captainraju1Bot; CaraBot; CarnivorousBot; CarsracBot; Cartridge-bot; CasteloBot; Castibot 92; CAT:TEMP deletion bot; CAT:TEMP deletion bot ; CataBotTsirel; CategorizationBot; CategoryBot; Category-bot; CategoryBot~enwiki; CategoryWatchlistBot; Caypartisbot; CbmBOT; CekPotBot; Cellistbot; CEM-bot; CensusBot; CERabot; Cerabot~enwiki; CerebrumBot; Cewbot; CG-bot; Chabbot; ChandlerMapBot; Chartbot; ChayitaBOT; Cheers!-bot; Chem-awb; CheMoBot; ChenzwBot; Cherybot; ChessBOT; CheWikibot; Chicobot; Chikamichi; Chlewbot; Chobot; Choboty; ChoreBot; ChoRusBot; Chris G Bot; Chris G Bot 2; Chris G Bot 3; Chrisbot; ChrisGBot2; ChrisSalij Bot; ChristianBot; ChuispastonBot; Chuli; ChzzBot; ChzzBot II; ChzzBot II ; ChzzBot II/header; ChzzBot III; ChzzBot III/header; ChzzBot IV; ChzzBot IV/header; ChzzBot V; ChzzBot V/header; CinemaBot; CipherBot; Citation bot; Citation bot 0; Citation bot 0 ; Citation bot 1; Citation bot 2; Citation bot 3; Citation bot 4; Citation bot test; Citation bot test ; CitationCleanerBot; CitationTool; CiteFixBot; Citing Bot; CivviBot; CJBot; ClanBot1; CleanBot; CleanupBot; CleanupBot~enwiki; CleanupListingBot; CleanupWorklistBot; ClerkBot; ClickBot; ClickBot (usurped); ClickBot (usurped) ; CloudNineBot; CLT20RecordsUpdateBot; ClueBot; ClueBot Commons; ClueBot Commons/Userpage; ClueBot II; ClueBot III; ClueBot IV; ClueBot IX; ClueBot NG; ClueBot V; ClueBot VI; ClueBot VII; ClueBot VIII; ClueBot VIII ; ClueBot X; ClueBotII; ClueBotIII; ClueBotIV; ClueBotVI; CmdrObot; CMoonBot; CobainBot; COBot; CobraBot; CocuBot; CodeMonkBot; CoderSIBot; CohesionBot; COIBot; CollabBot; CollabRCBot; Collabsearch; Colourful Bot; Comics-awb; CommaBot; Commander Keane bot; Commander Keane bot ; Commonb0t; Commons fair use upload bot; Commons fair use upload bot ; Commons fair use upload bot (usurped); CommonsBot; CommonsDelinker; CommonsDelinquent; CommonsNotification; CommonsNotificationBot; Community Tech bot; Compteur d'_ditions (bot); Computermacgyverbot; ConnectiBot; ContentCreationBOT; ContinuityBot; CoolMaster12345; CopperBot; CopyToWiktionaryBot; CorenANIBot; CorenBlockMonBot; CorenProxyBot; CorenSearchBot; Coreva-Bot; Correcteur de redirection; CosineBot; CounterVandalismBot; CountryBot; Courcelles Bot; Courcelles Bot ; CpiralBot; Cprobot; Crash bot; CrazynasBot; CrazyPhunkbot; CreaBOT; CricketBot; CrimsonBot; CrniBot; Crochet.david.bot; Cronbot; CruccoBot; Crypticbot; CSBot; CSDCheckBot; CSDWarnBot; CsurlaBot; CTMakerbot; CultureBot; CurlyBot; Curpsbot-unicodify; CurtaintoadBOT; CVBOT; CVNBot; CwengerBot; CwraschkeDataBot; Cwt96bot; CX42 Bot; Cyberbot; Cyberbot I; Cyberbot II; Cyberbot III; Cyberbot Trial Bot; CyberbotII; Cydebot; CyeZBot; CyroBot; CyroBot~enwiki; CzechoBot; D6; DabBot; DaimonBot; DalcherBot; DallasBot; DaneBot; DanhashBot; DanielBot; Danielfolsom2.bot; Danilo.bot; DanjanBot; DanKoehlBot; DanmicholoBot; Danroks bot; Dante Defrenden; Danumber1bot; DarafshBot; DarioBot; DarioBot~enwiki; Dark Shikari Bot; Dark Shikari Bot ; Darkicebot; Darkicebot II; Darkicebot II ; DarknessBot; DarwinBot; DasBot; DASHBot; DASHBotAV; Dasschaefchenbot; DastyorBot; DatabaseBot; DataflowBot; DatBot; Dateientlinkerbot; DavepapeBot; DavidLeighEllisBot; DavidWSBot; DbBot; Dbl2010bot; Dcirovicbot; DcoetzeeBot; DcoetzeeBot~enwiki; DeadBot; DeadLinkBOT; DEagleBot; Dedebot; DeepBot; DefaultsortBot; DefconBot; DeleteAsstBot; DeliveryBot; DeltaQuadBot; Demibot; DemocraticsBot; Denbot; DeniBot; DennisPeetersBot; DeprecatedFixerBot; DerbethBot; DermBOT; Descriptioncreator; Detroiterbot; Dexbot; DFBot; DhanakBot; DHN-bot; DHN-bot~enwiki; Dibot; DickensBot; DieBucheBot; Diego Grez Bot; DiegoGrezBot; DigitalmeBot; DiliBot; Diligent Terrier Bot; Diligent Terrier Bot ; Dillonbot; DimaBot; Dinamik-bot; DinoBot; DinoBot2; DinojermBot; DinoSizedBot; Dinybot; DipankanBot; DirlBot; DisambigRedirBot; Disambot; DividingBot; DixonDBot; DiyarBot; Djobot; Dl2000; DMbotY; Doddebot; DodekBot; DodekBot~frwiki; DodoBot; D'ohBot; DOI bot; Dokomonta-Bot; DomBot; DomBot~frwiki; DonabelSDSU.bot; DonnerJack.bot; Doombot; DorganBot; DottyQuoteBot; DougBot; DownloadBot; DoyleyBot; DP Bot; DP Bot ; Dpkbot; DPL bot; DPLbot; DpmukBOT; Dr Bot; DragonBot; DragonBot~pmswiki; DraiconeBot; DreamBot; Dreamy Jazz Bot; DrFO.Tn.Bot; DrilBot; Drinibot; DRN clerk bot; Drng-bot; DrTrigonBot; DRVBot; DschwenBot; DSisyphBot; DuckBot; Dudubot; DumbBOT; DumZiBoT; DustaBot; DusterBot; DustyBot; DvorapaBot; DvyBot; DyceBot; DYKadminBot; DYKBot; DYKHousekeepingBot; DYKReviewBot; DYKUpdateBot; Dylan620 Bot; Dylan620 Bot ; DysklyverBOT; E2mb0t; EagleAstroidBot; EaglesBot; EarwigBot; EarwigBot I; EarwigBot III; EastEnders Bot; EberBot; EBot; EBot II; EBot II ; EBot III; EBot III ; EBot IV; EBot IV ; Ebrambot; EchoBot; EdBot; Edbot~enwiki; EddieBot; EdgarsBot; EdinBot; Edmundobot; EdoBot; Edoderoobot; EdwardsBot; Edwardspec TalkBot; EFixBot; Eflybot; Egmontbot; EgressBot; EinsBot; Einsteinbot; Eivindbot; EjsBot; El bot de la dieta; El bot de la dieta ; EleferenBot; Elessar; Elfobot; Elhuyar Fundazioa; ElimBot; Elissonbot; EllisBot; ElMeBot; ElphiBot; Elvisor; Email Bot; EmailBot; EmausBot; EmBOTellado; Emijrpbot; EmptyBot; EmxBot; EnDecoderBot; ENewsBot; Englishbot; EnigmaBot; Enrichyourmind; Ent~frwiki; EnterpriseyBot; EnzaiBot; EnzetBot; EnzoBot; EpopBot; Erabot; EranBot; ErfgoedBot; Erik E VestBot; Erik E VestBot ; Erik9bot; ERJANIKbot; Erwin85Bot; Erwin85TBot; Erwin-Bot; ESBot; Escarbot; Eskimbot; Eskimospy Bot; Eskimospy Bot ; EsquivalienceBot; EssjayBot; EssjayBot II; EssjayBot II ; EssjayBot III; EssjayBot III ; EssjayBot IV; EssjayBot IV ; Estirabot; Ethen12bot; Ethen12bot 2; Ethen12bot 2 ; Etherbot; EthicalBot; Eubot; EukeshBot; EuseBot; EVA (bot); EVA2.0 (bot); Example Bot; Example Bot ; ExodusBot; ExpertCommentBot; ExpertIdeasBot; ExpireBot; Eybot; Eybot~enwiki; EyeBot; EyeEightDestroyerBot; FA RotBot; FA Template Protection Bot; FA Template Protection Bot ; FA2010Bot; FACBot; FacebookBot; FahBot; Fair-Use Bot; Fair-Use Bot ; FairuseBot; Fajrbot; FaleBot; FameownerBot; FANSTARbot; FarbodBot; FariBot; FastilyBot; Fatemibot; Fatranslator; Fawikibot; Fbot; FDA to Wikipedia Bot; FearBot; FeedBot; Felagund-bot; FelixBot; Femto Bot; Fetofsbot; Fetofsbot2; Fettgesicht; Ficbot; Filbot; File Upload Bot (AntWeb); File Upload Bot (Kaldari); File Upload Bot (Kaldari) ; FileBot; Filedelinkerbot; FilRBot; FiriBot; FischBot; Fixatypobot; FixBot; FixDateBot; FixyBot; FlaBot; FlagBot; flBot; Flickr upload bot; FlickreviewR; FloBot; FloggerBot; Flow talk page manager; Fluffbot; FlutefluteBot; Fluxbot;

FMAFanBot; FMTbot; FNBot; FoBeBot; Fobox; FolksourceBot; FotothekBot; FoxBot; FPBot; FraBOT; FractalBot; FRadical Bot; Framabot; Francosrodriguez; FrescoBot; Frettiebot; Fridae'sDoomBot; Fritzbot; FritzpollBot; FroggyBot; FSBot I; FSBot I ; Full-date unlinking bot; FWTestBot; G.bot; G_T; G3robot; GA bot; Gabriel-Bot; Gabrielchihonglee-Bot; Gacbot; GaemariBot; GaiJinBot; Gallicbot; Galobot; GameOnBot; GameSpot-Eng-Bot; Ganeshbot; GargoyleBot; GBenemyBot; G-Bot; G-Bot~enwiki; GCIBot; Gdr/DYKbot; Gdr/Nomialbot; Gdr/Yearbot; Gdrbot; Geimas5Bot; Geimas5Bot~enwiki; Genderlessbot; GeneratorBot; GeniusBot; Geobot; GeographBot; GeorgeMoneyBot; GeorgeMoneyBot-status; Gerakibot; Gerbot; GerWsUpload; Getbot; GhalyBot; GhosterBot; GIBBot; GiftBot; Giggabot; GilerBot; GimmeBot; Ginosbot; Giro bot; GMDBot; GnawnBot; Gnome ; Gnome (Bot); Gnome (Bot) ; GnuBotmarcoo; Goalkeeperbot; GoatBot; GoblinBot4; Gobot5555; GoeBOThe; GondranBot; Good Article Patrol Bot; Good Article Patrol Bot ; GPHbot; Gpvosbot; GracenotesBot; Grafikbot; Grammarbot; GrashoofdBot; GrassnBreadBot; GrassnBreadBot~enwiki; Great Bot on Fire; GreenC bot; GreenC bot/Job 5; GreenCbot; GregBot; Grillbot; Grillitus; Grim Reaper Bot; Grim Reaper Bot ; GrinBot; GrinBot~enwiki; GrooveBot; GrouchoBot; GTBot; Guanabot; Guanabot~frwiki; Guanabot2; Guardian Robot 3568NG; Guardian Robot 3568NG ; Gubot; Guoguo12Bot; GurchBot; GyaBot; GZ-Bot; Gzen92Bot; H.b.shbot; H2Bot; H3llBot; H92Bot; HagermanBot; HairBot; Halibott; Hamlet Prince of Robots; Hamlet Prince of Robots ; HangsnaBot; HanhilBot; HappyBot; HaqmarBot; Harej bot; HarrivBOT; Harrobot; Hashar; HasharBot; HasharBot~enwiki; HasteurBot; HawkEyeBot; Hawk-Eye-Bot; Hazard-Bot; HBC AIV helperbot; HBC AIV helperbot 10; HBC AIV helperbot 8; HBC AIV helperbot 8 ; HBC AIV helperbot11; HBC AIV helperbot12; HBC AIV helperbot2; HBC AIV helperbot3; HBC AIV helperbot4; HBC AIV helperbot4 ; HBC AIV helperbot5; HBC AIV helperbot7; HBC AIV helperbot9; HBC archive builderbot; HBC archive builderbot ; HBC Archive Indexerbot; HBC Archive Indexerbot ; HBC Archive Indexerbot/OptIn; HBC NameWatcherBot; HBC RenameClerkBot; HBC RenameClerkBot ; HBCAIVhelperbot; HBCAIVhelperbot2; HBCAIVhelperbot3; HBCAIVhelperbot8; HBCNameWatcherBot; H-Bot~fiwiki; HDstars.bot; HedgeBot; Heikobot; Helpful Pixie Bot; HelpfulPixieBot; Helsabot; HeltBot; Henry McCleanBot; HerculeBot; HersfoldArbClerkBot; HersfoldBot; HersfoldCiteBot; Hexabot; Hgzbot; HhhipBot; HiDrNickBot; Hintssbot; HiTeCBot; HiW-Bot; Hkbot; Hkn-bot; Hmainsbot; Hmainsbot1; HMBot; HMBot~enwiki; Hockeybot; Holek.Bot; Homobot; HooftBot; Horst Fuchs; HostBot; HotArticlesBot; HRoestBot; HsfBot; HtonlBot; HUBOT; HueSatLumBot; Hugo999; HujiBot; Humbot; HunsuBot; HunyadymBot; Huzzlet the bot; Huzzlet the bot ; Hxhbot; HydraBot; HypoBOT; HyuBoT; I,Robot; IagaBot; IainadamsBot; IanraBot; IByte/Bot; IcalaniseBot; IDbot; IdeoBot; Idioma-bot; Ilmari Karonen's adminbot; Ilmari Karonen's adminbot ; IluvatarBot; Image Rapture Bot; Image Rapture Bot ; Image Screening Bot; ImageBacklogBot; ImageBot; ImageRemovalBot; Image-req-proj-bot; ImageResizeBot; ImageScreeningBot; ImageTagBot; ImageTaggingBot; Imbot1; InactivityEmailBot; InceptionBot; IncolaBot; IndvTbot; IndyJrBot; InfoBot; InfoBot~enwiki; Infobox Bot; Infobox Bot ; Infobox Journal Update Bot; Infobox Journal Update Bot ; Infobox Wikipedia bot; Infobox Wikipedia bot/doc; Infobox Wikipedia bot/sandbox; InfoboxBot; InkoBot; Innocent bot; Innocent bot ; Innocent datumbot; Innocent iwbot; Innocent iwbot ; InsideoutBot; InstructorCommentBot; Int21hBot; InternetArchiveBot; InterwikiConversionBot; Interwikis; Invadibot; IonutzmovieBot; Ipatrol-bot; IPLRecordsUpdateBot; IPTaggerBot; Ir4ubot; Iradigalesc Bot; Irfan-bot; IronWarrior EditBot; IsraBot; Iswatchbot; Italic title bot; Itubot; IvanBot; IwaimBot; IW-BOT; IznoBot; J Milburn Bot; J Milburn Bot ; JabbaTheBot; Jack Rabbot; JackBot; JackieBot; JaGaBot; Jager Bot; Jager Bot ; Jager Bot/Userpage; JagRoBot; JaguarBot; Jakebot; JamietwBot; JAnDbot; Janna Isabot; Janna Isabot ; Japbot; Japiobot; JarBot; JarektBot; JaskaBOT; JatBot; Jayden54Bot; Jbawt; JBradley Bot; JCarvalhoBot; JCbot; JCbot 2; JCbot 2 ; JCW-CleanerBot; JdforresterBot; Jeblad (bot); JeffGBot; Jeremybot; Jerodlycett-autobot; Jeroenbot; JerryBot; JerryBot/FAQ; JhealdBot; JhsBot; Jimmy-bot; Jitse's bot; JJBot; JJMC89 bot; JJMC89 bot II; JL-Bot; Jmax-bot; JMuniBot; Jnanabot; JoaoMirandaBot; JobuBot; JoeBot; JoelHelperBot; Joel's Bot; Joe's Null Bot; Joe's Olympic Bot; Jogersbot; JogoBot; John Bot; John Bot ; John Bot II ; John Bot III; John Bot/Header; John of Reading Bot; Johnbot; JohnFLBot; JonathanBot; Jonjonbot; Jonny-bot; Joopwikibot; JoRobot; Josh3580 BOT; Josh3580 BOT ; JoshurBot; Josvebot; Jotterbot; Jozef-k.bot; Jpbot; Jsgray1993; JThorneBOT; JThorneBOT/sandbox; Jumbuck; Justincheng12345-bot; JVbot; Jwbot; JYBot; JYBot/monobook.css; K.bot; KadaneBot; Kaiger; KaiserbBot; Kakashi Bot; KakashiBot; KaldariBot; Kal-El-Bot; KamikazeBot; Kanjybot; KarBOT; KarlsenBot; KartikBot; kasirbot; KasparBot; Kaspobot; KatBot; Kbdankbot; Ken123BOT; Kenrick95Bot; Kerberizer; KevinalewisBot; KevinBot; Kgsbot; Khabot; KhanBot; KharBot; KhiviBot; KhrisBot; KhunterBot; Khutuck Bot; KidsBot; KielimiliisiBot; Kikobot; KiloBot; Kingbotk; KingpinBot; KingRbot; Kiril-Bot; Kirkbot; Kisbesbot; KITbot; KittyBot; KiwiBot; Kl4m-AWB; KLBot2; KMLbot; Knedlik-Pod; KnightRider; KnightRider~enwiki; KnopfBot; Kobotbel; KocjoBot; KocjoBot~enwiki; KoehlBot; KolbertBot; KolBot; Kolega2357-Bot; Kotbot; KotzBot; KrattBot; KrBot; Krdbot; KrimpBot; KRLS Bot; KSBot; KSFT bot; KslotteBot; KuBOT; KuduBot; Kumar Appaiah Bot; Kumar Appaiah Bot ; Kumi-Taskbot; KumulBot; Kungfubot; KunMilanoRobot; Kurando-san; Kururubot; Kwjbot; Kyle the bot; Kyle the bot ; KyluBot; L PBot; L&K-Bot; L.A. (AWB); L293D (AWB); LA2-bot; LaaknorBot; Lady Akasha Bot; Lait ribot; LambdaBot; Langbot; Lang-Bot-as; Langtoolbot; LanguageBot; LaninBot; LankLinkBot; LantayBot; LaraBot; LarBot; LarskeBot; Latexbot; LatitudeBot; LauBot; LawBot; Lcarsbot; LDBot; Le Chat BotT_; Le Pied-bot; Le Pied-bot~enwiki; Le plus bot; LeadSongDog/Infobox bot sandbox; Lechat bot; LeChatBot; Legobot; Legobot II; Legobot II ; Legobot III; LeoBot; LeonardoRob0t; Leperebot; LerdsuwaBot; Liangent-bot; Lightbot; Ligulembot; LijeBot; LimitedWP:CANVASandCleaningupafterbotsyntaxerrors; LinedBot; LinkFA-Bot; LioneltBot; Liquid-aim-bot; Liso-Bot; ListasBot; ListeriaBot; ListGenBot; ListManBot; Livetslott; Livetslotteri; LivingBot; Llamabot; Lockalbot; Locke Bot; Locke Bot ; Locobot; LoganBot; LolBot; LOLibot; LolsimonBot; Longbot; Lonjers french region rename bot; LoquBot; LordAnubisBOT; LordBumbury; LostBot; Lot-bot-as; Louperibot; LourdesBot; LoveBot; Loveless; Lowercase sigmabot; Lowercase sigmabot I; Lowercase sigmabot II; Lowercase sigmabot III; Lowercase sigmabot IV; Lowercase sigmabot V; LSG1-Bot; Lsjbot; Lt-wiki-bot; Lt-wiki-bot~frwiki; LuaBot; Luanbot; Luas g bot; LuasÊg bot; LuasÊgbot; Luasóg bot; Lucasbfrbot; Lucia Bot; LucienBOT; Luckas-bot; LuddsBot; Ludo29bot; LuisBot01; Luke081515Bot; Luki-Bot; Lunabot; LupinBot; Luuvabot; Lvp-bot; LymaBot; LyricsBot; M4x0uBot; M7bot; Maathavanbot; MaatyBot; Mabdulbot; Mabot; MacMed auto; MacMed auto ; MacMedBot; Macro Bot; Macro Bot ; Madhubot; MadmanBot; MadMarkbot; MadPrav; Maelgwnbot; Magic links bot; MagnusA.Bot; MagulBot; MahdiBot; Main Page Image Bot; Main Page Image Bot ; Maintenance script; Mainulmizan bot; Mairibot; MajedBot; MajordomoBot; MakBot; Makecat-bot; Maksim-bot; MalafayaBot; malarzBOT; MammothBot; MandelBot; Manishbot; Manubot; MaraBot; Marbott; MarcBot; Marco27Bot; Margosbot; Margosbot~enwiki; MarkkoBot; MarmaseBot; MarshBot; MartinBot; MartinBot4647; MartinBotII; MartinBotIII; MartinBotIV; Martin's bot; Martin'sbot; Maskbot; Mason1213; MasterMatt-Bot; MastersBot; mastiBot; Matanyabot; Matbot85; Mathbot; MathoBot; MatmaBot; MatSuBot; Matt Crypto (bot); Matt Crypto (bot) ; Matthewrbot; Matthewrbowker/WikiWelcomer; MauchoBot; MaundBot; MauritsBot; MayorofrosharonBot; Mayurbot; MazinBot; MBHbot; MBisanzBot; MBot; M-Bot; MBot~enwiki; MCBot; McM.bot; Mcrbot; Mdann52 bot; Mdann52 IMG bot; Me iwan; MedBot; MedcabBot; MedcabBot2; MedHrox; MediationBot; MediationBot1; MediaWiki default; MediaWiki message delivery; MeerderBot; Meker

Edtitor1; Melancholie; MelancholieBot; MelonBot; Melouresbot; Memty Bot; Memty Bot ; MenasimBot; Mendeleevbot; MenoBot; MenoBot II; Mentibot; Merge; Merge bot; MerlBot; MerlIwBot; MerlLinkBot; Mersenbot; MessageDeliveryBot; MessedRobot; MessengerBot; MessengerBot~enwiki; MetaplasticityBot; MetapointBot; Metriki; MetrikiBot; MetroBot; MetsBot; MexicanoBot; MfDBot; MGA73bot; MGA73bot2; Mgmbot; MGriBot; MgSunBot; MHP07Bot; Micbot; MichaelBillingtonBot; MichaelkourlasBot; Michaello; MickiBot; MifterBot; MifterBot I; MifterBotI; Miguillen-bot; MihaitzaBot; Mihas-bot; MikaniBot; MikemoralBot; MilHistBot; MilicevicBot; Milk's Favorite Bot; Milk's Favorite Bot ; Milk's Favorite Bot II; Milk's Favorite Bot II ; Milk's Favorite Bot III; MimihitamBot; Mindlramibot; Minsbot; MinusBot; Mir Almaat 1 S1/Aleppobot; MirgolthBot; MiszaBot; MiszaBot I; MiszaBot II; MiszaBot III; MiszaBotIII; MixBot; Mjbmrbot; Mkdw Bot; Mkdw Bot ; MMABot; Mmbot; MMOGMailBot; MMOGMailMan; MnidBot; MobyBot; MohandesBot; MolandBot; MomijiRoBot; MondalorBot; Monkbot; monobook.css; MonoBot; Montibot; Montibot/sandbox; MoohanBOT; MoondyneAWB; MoondyneBot; MoovidaBot; MorbZ-Bot; MOSNUM Bot; MotinBot; MotnahpBot; Movses-bot; Mpradeepbot; MPUploadBot; Mr.Ibrahembot; Mr.Z-bot; MrBlueBot; MrVanBot; MSBOT; MTC RoBOT; MTSbot; Muhaiminkudou; Mulder416sBot; Muninnbot; Munjabot~usurped; Muro Bot; Muro Bot ; MusikBot; MusikBot II; MusikyanBot; Mutleybot; MuZebot; MyBot; MystBot; N96bot; NaggoBot; NahidSultanBot; Nal-Bot; Nallimbot; NameBot; NameBot~enwiki; Namobot; NanoBot; Naohiro19bot; NapalmBot; Naqvibot; Nasko.bot; NaskoRobot; NathanBot; Natig98bot; NationalRegisterBot; Naudefjbot; Naudefjbot~enwiki; Navibot; Navi-bot; N-Bot; Nearly Headless Bot; Nearly Headless Bot ; NedBot; NeechalBOT; NekoBot; NekoDaemon; Neobot; Neombot; NeraBot; NerdyScienceBot; NetBot; NeuRobot; Newsletterbot; Nguoimay; Nichalp-bot; NickyBot; NihiltresBot; NihlusBOT; Nikbot; NilfaBot; NinoBot; NirmosBot; Nixbot; NjardarBot; NjsBot; NNBot; NNBot II; NNBot II ; NobelBot; Noble Story Bot; Noble Story Bot ; Nobot; Nobot~enwiki; NoclaimsBot; NodBot; NohatBot; NolBot; Nomenclaturebrowser; Nomiff; Non-Free Content Compliance Bot; Non-Free Content Compliance Bot ; NongBot; Nono le petit robot; NoomBot; NorefBot; NotACoolBot; Notavailablex; Nothingbot; NovBot; NowCommons-Sichtbot; NrhpBot; NTBot; NuclearBot; NukeBot; Nullzerobot; Numbo3; Numbo3-bot; Numptybot; NVS; NVS(bot); NW557Bot; NyappyBOT; NyenyecBot; NyokiBot; Nyubot; Nyxbot; O bot; O bot ; OAbot; Obaid-bot; ObelixBot; Obersachsebot; Ochbot; OchilovBot; Ocobot; OctraBot; OdderBot; OfekBot; OgreBot; Ohms Law Bot; OhmsLawBot; OKBot; OKBot II ; OKBot II~enwiki; Olafbot; OldMedcabBot; OmarGhridaBot; OmniBot; One bot; ONKbot; OpenlibraryBot; Ops Monitor ; Ops Monitor (WMF); OrBot; Orbot1; Orgullobot; Orgullobot~eswiki; Orgullobot~frwiki; OrikriBot; OrlodrimBot; orofrosharonBot; OrophinBot; OrphanBot; Orphaned image deletion bot; Orphaned image deletion bot ; Orphaned talkpage deletion bot; OrtoBot; OverlordQBot; OverlordQBot/BRFA2; OwtbBot; OxBot; P31n3/sandbox; Page correction BOT; Pagecount Bot; Pagecount Bot ; Pageview bot; Pageview bot ; PaievBot; PaintBot; Palica; PalicaBOT; PalnaBot; PALZ9000; Panicbot; PanzerBot; Papa6-bot; ParaBot I; ParaBot I ; Parent5446 Bot; Parent5446 Bot ; PasabaUnBotPorAqui; PascalBot; Pasquebot; PastoriBot; Pasztillabot; Patator Bot; PatersBot; Pathbot; Pathosbot; Patrick87-Bot; PatruBOT; Pattonbot; PattyBot; Paulatz bot; PavloChemBot; Pawe_ Ziemian BOT; PbBot; PBot; P-bot; PC78-bot; PCbot; PCM; PDBbot; PDFbot; PearlBot; Pearle; Peelbot; PeerReviewBot; Pegasusbot; PenguinBot; People-n-photo-bot; People-photo-bot; Perebot; Perebot~enwiki; Perebot~frwiki; PerlegoEditorBot; PertBot; Petan-Bot; Peter17-Bot; Peti610bot; Peti610botH; Petr Kyborg; Petr Kyborg ; Peykbot; PfamWikiBot; Pfft Bot; Pfft Bot ; Pfft Bot~enwiki; Phe-bot; PhiloBot; Philosobot; Philosopher-Bot; Phoenix-bot; PhotoCatBot; PhuzBot; PhyloBot; Pi bot; PidlisnukBot; PieRRoBoT; PikminBot; Pil56-bot; Ping08Bot; PinoBot; PinpointBot; PipepBot; PiRSquared17Bot; Pi's Bot; Pi's Pixie Bot; PityuBot; PixelBot; PkbwcgsBot; PlacenamesBot; PladaskBot; PlangeBot; PlankBot; Planktonbot; Plasticbot; PleaseStand (bot); PleaseStand (bot) ; PmegBot; PNG crusade bot; PNG crusade bot ; PNG recompression; PNG recompression ; PockBot; PocKleanBot; PointBot; Pokbot; PolarBot; Polbot; Polbot/older tasks; Pompidombot; PoplLyricsBot; Porchcorpterbot; Portal box bot; PortalBot; PortalBot~enwiki; Porthos; PorthosBot; PostBot; PostOnTalk; PotatoBot; PowerBot; PoxBot; Pparazorbot; Pp-bot; Prabot; PraetorianCheese; Pranbot; Prebot; PrimeBOT; ProcseeBot; Prof.Bot.1; Profbot; ProgrammingBot; ProgreSS; Project Messenger Bot; Project Rastko bot; Project Rastko bot ; ProjectBot; ProjectMessengerBot; ProjectRequestedPagesBot; Prom_th_eBot; Prombot; ProtectionBot; ProtectionTaggingBot; ProteinBoxBot; ProtLinkbot; Proxybot; PrzemuBot; PsBot; PseudoBot; PsychAWB; Pszcz†_ka; Ptbotgourou; Puggansbot; PUNG BOT; Purbo T; PurboT; Putorobot; PvssBot; PxBot; PxBot II; PxBot II ; QBA-bot; Qbot; Qbugbot; QOTDBot; QovulwBot; QstBot; QualiaBot; Quark-bot; Quarrybot-enwiki; QuerubBOT; QueryBot; QuestionCopyright; Quotebot; R Delivery Bot; R Delivery Bot ; R. Hillgentleman; R. Hillgentleman ; R. Koot (bot); R. Koot (bot) ; R• bot ; R28Bot; Rabbot; RaBOTnik; Rachmat-bot; RadufanBot; RaftaarBot; RagesossBot; RagibBot; Rahhib; RahmanuddinBot; Rain night-AWB; Rainbot; RajeshBot; Ralbot; Ramaksoud2000Bot; Rambot; Rameshngbot; RamissesBot; Rangerbot; RanZbot; RapidBot; RaptureBot; RarBot; RavpawliszBot; Razibot; RBot; RBot~enwiki; RBSpamAnalyzerBot; RCBot; RCBot~enwiki; RCI~enwiki; Reader bot; Reader bot ; ReaperBot; ReaperBot X; ReaperBot X ; RebelRobot; Red Thrush's Bot; Red Thrush's Bot ; RedBot; Redirect fixer; Reedy Bot; ReedyBot; RefBot; RefDeskBot; ReferenceBot; RefSpaceBot; Region102Bot; Rei-bot; ReigneBOT; ReinaartBot; RelistBot; Rembiapo pohyiete (bot); RemindMeBot; Renabot; RepairBot; Reports bot; Revert-Statistik; Rezabot; Rfambot; RFC bot; RFC bot ; RFC posting script; RFF-Bot; RFRBot; RHbot; Riad.Bot; Riad.Bot~enwiki; RibotBOT; Riccardobot; RichardcavellBot; Rick Bot; Rick Bot/test; RileyBot; ringobot; Ripchip Bot; Ripebot; Rjbot; RjpBot; RjwilmsiBot; RKBot; RLutsBot; RM bot; RMCD bot; Rø bot; Rob110178bot; Robbot; RobbyBot; Robchurch/Bot; Robert SkyBot; RobinBot; RobinHood70Bot; Robo5; Roboconvallaria; RoboDick; RoboDick~enwiki; Robodoc.at-fr; RobokoBot; RoboMaxCyberSem; RoboServien; robot; Robot du 90; RobotE; RobotG; Roboti Tung; Robotic Garden; Robotic Garden ; Robotito; RobotJcb; RobotMichiel1972; Roboto de Ajvol; Roboto de Ajvol ; Robotpjetter; RobotQuistnix; RobotSC; RobotTbc; Robthbot; RocketBot; RockfangBot; Rodrigo Padula (BOT); Rodrigolopesbot; RoggBot; Roland45-Bot; RollbackerBOT; RolloBot; RomaineBot; RomanianFootballBot; RomanianFootballBot/Userboxes; RomulusBot; RonaldBot; RonBot; RooBot; Roomba; Rootology Bot; Rootology Bot ; RoryBot; Rotbot; RotlinkBot; Rovibot; Rpabot; Rschen7754bot; RscprinterBot; RSElectionBot; Rtz-bot; Rubinbot; RudolphousBot; RussBot; RWHbot; Ryan Vesey Bot; Ryan-Bot; RyBot I; RyBot I ; RyRyBot; S205643bot; Saadkhan12345.bot; Safibot; Sagabot; SagaCookBot; SageBot; Sahimrobot; Salebot; Salebot III; SalebotJunior; SalviBot; SamatBot; Sambot; SAMI.Bot; SamoaBot; SamuraiBot; Sandbot; SandgemBot; Sanjeev bot; SanniBot; Sanolnacobot; SantoshBot; Sarojbot; SashatoBot; SashkoBot; SassoBot; SassoBot/Header; SassoBot/Status; Satdeepbot; SatyrBot; Sayyid Azzam Mufhadol; ScanDot; Scepbot; SCG01Bot; SCGBot; SchlurcherBot; SchuBota; ScoopBot; ScottLantz2; ScriptNastaliqBot; Scsbot; SD5bot; SdBot; Sdebot; SDPatrolBot; SDPatrolBot II; SDPatrolBotII; Sebbot; SebrevBOT; SecuniBot; Seedbot; Seedbot2; SelectionBot; SelketBot; Selmobot; Sembot; SergoBot; Sethbot; SeventyThreeBot; SeveroBot; SEWilcoBot; Sgeo-BOT; SgtDrini; ShabBot; Shadowbot; Shadowbot2; Shadowbot3; ShadowRobot; ShakataBot; ShakingBot; Shanbot; SharafBot; SharedIPArchiveBot; Shareobot; SharkDBot; SHBot; SheepBot; ShepBot; ShepBot/Info; Sherpa; ShinBot; ShinmaBot; Shirobot; ShitiBot; Shoutboundbot; ShrugBot; Shuaib-bot; Siddhartha Ghai bot; SiddiqBot; SidoBot; SieBot; Sigmabot; Signpost Book Bot; SignpostBookBot; SilvonenBot; SimoneBot; Simplebot;

SimplexityBot; SineBot; Sir Louis Point du Lac Bot; SireBot; SJSF-2-Bot; SJSFBot; Sjundebot; SKbot; Sk-Bot; SkiersBot; Skumarlabot; Skxll Mark-2; SkylinBot; SlakrBot; SLMSBot; Slobot; Slowbot; SLQbot; SmackallBot; SmackBot; Smallbot; Smartbot; SmeiraBot; SmeiraBot~enwiki; SmileBot; SMS Bot; SMS Bot ; SMS Bot 2; SMS Bot 2 ; Sn1pebot; Sn1pebot Mk. II; Snaevar-bot; Snotbot; Snowbot; SoapBot; Soccoro; SocialBot; SocietyBot; SocksBot; SolumeirasBot; SOM TELL HELPER; SOM TELL HELPER/Task; SOMBot; So-Mr.Z-bot; SoraBot; Soulbot; SoupBot; source codezzz; SourceArchiver; SoxBot; SoxBot II; SoxBot III; SoxBot IV; SoxBot IX; SoxBot V; SoxBot VI; SoxBot VII; SoxBot VIII; SoxBot VIII ; SoxBot X; SoxrockBot; SpaceFactsBot; SpadeBot; SpammerTrapBot; SpamReportBot; SpamReportBot~enwiki; SpBot; SPCUClerkbot; SpeakerBot; SpebiBot; Speedybot; SpeedyGonzalez; Spelian; SpellCheckerBot; SpellingBot; SpillingBot; SplineBot; SpongeBot (usurped); SpongeBot (usurped) ; SporkBot; SportiBot; SportsStatsBot; SPPatrolBot; SpreadthesignBot; SprinterBot; SQLBot; SQLBot-Hello; SreeBot; SSTbot; Staeckerbot; StanfordLinkBot; StanfordLinkPredictor; StarusBot; StatisticianBot; StatsBot; StatusBot; StatusBot/Status/Fercho85; STBot; STBotD; STBotI; STBotT; Steenthbot; SteenthIWbot; Stefan2bot; StefanBot; StefBot; SteveBot; StigBot; StillwaterBot; StormBot; StradBot; Strainubot; Structor; STTWbot; Stuartyeates randombot; StubListBot; StubSyncBot; Stv.bot; Stwalkerbot; STymBot; SubstBot; SubstBot~enwiki; SudokuBing; Sudoplusii; SuggestBot; SuisuiBot; SUL Bot; SUL Bot ; Sumanthk; Sumibot; Sumone's bot; SunBot; SunCreatorBot; SundarBot; SupBot; SuperBot; Superninobot; Superzerocool; SuprememangaBot; SurenaBot; SusBot; SusBot~enwiki; Svenbot; Svensson1bot; SVGBot; SvickBOT; SVnaGBot1; SW9Bot; SwiftBot; Swimmingbot-awb; Swindbot; SynBot; SyntaxTerrorBot; Synthbot; Synthebot; Syrcatbot; Szczepan.bot; Sz-iwbot; T13bot; T-850 Robotic Assistant; T-850 Robotic Assistant ; TabellenBot; TaBOT-zerem; Tadas12-bot; Tadasubot; TaeBot; Tahir-bot; Talabot; Tangobot; Tanhabot; TannerBot; Tanqbot; TAP Bot; TARBOT; TarHippoBot; TaskForceBot; Tatobot; TauerBot; Tawbot; Tawbot~enwiki; Tawkerbot; Tawkerbot2; Tawkerbot4; TawkerbotTorA; Taxelbot; Taxobot; Taxobot ; Taxobot (usurped); Taxobot 1; Taxobot 1 ; Taxobot 2; Taxobot 3; Taxobot 6; Taxobot 7; TaxonBot; TaxonKatBot; Tcho_oBot; TecBot; TechBot; TechniqueBot; Technophant/Archive 1; TedBot; TedderBot; Tegebot; TekBot; TemirovBot; TEMPbot; Template Maintenance Bot; Template Maintenance Bot ; TemplateBot; TennisBot; TeoBot; Terrabot; TeslaBot; TestEditBot; Test-tools~frwiki; TethBot; TextworkerBot; Texvc2LaTeXBot; TFA Protector Bot; TFAProtectorBot; Thadius856AWB; Thanatos bot; ThBot; The Anomebot ; The Anomebot2; The Anonybot; The Auto-categorizing Robot; The Little Pixie's Friend; The Mini PEKKA; The Polish Bot; The RedBot; The wubbot; The wubbot ; TheAuto-categorizingRobot; Thedjbot; Thehelpfulbot; TheJoshBot; TheklanBot; TheMagikBOT; Theo's Little Bot; Theo's Little Bot II; ThePhantomBot; Theroomen; TheSandBot; ThetaBot; Thijs!bot; Thijssie!bot; ThorBot; ThreeBot; ThundaBot; TianyammBOT; Tigraan-testbot; Tildebot; TinucherianBot; TinucherianBot II; TinucherianBot II ; TinucherianBot III; TiriBOT; TjBot; TLAbot; TobeBot; ToBot; ToePeu bot; ToePeu.bot; ToffileBot; TohaomgBot; TokenzeroBot; Tokrkbot; Tom.Bot; Tombot; TomcsyBot; Tom's Tagging Bot; Tomtomn00 Bot I; TonyBot; ToolWikiBot; Topjabot; Topjabot~enwiki; TorBlockingScript; TorBot; TorNodeBot; TortoiseBot; TottyBot; Tourjouman05Bot; TowBot; Town-bot; TPBot; TPO-bot; TptBot; TraBot; Tradubot; Trailbehind-bot; Traveler100/Photography-Proj-bot; TrBot; Tribot; TrionaBot; TronaBot; TronBot; TrustMeImAIRobot; TrustNoBot; Tsca.bot; TSM Fairuse Bot; TSM Fairuse Bot ; Tsunderebot; TTObot; TuanminhBot; TuanUt-Bot!; TuHan-Bot; Tullbot; Tullbot~enwiki; Tulsibot; Tuonela; Turk_szBot; TuvicBot; TweetCiteBot; TWLBot; TwsxBot; TXiKiBoT; TyAbot; Typebot; TypoBot; UAABot; U-bot; UcuchaBot; UgenBot; Ugur Basak Bot; Ugur Basak Bot ; Ugur Basak Bot~enwiki; UK Legislation Bot; UK Legislation Bot ; UltraBot; Umarbot; UnCatBot; Un

References:

Shalizi, C. R., & Klinkner, K. L. (2004). Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences. Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference (UAI 2004), 504-511.

Darmon, D. (2015). Statistical Methods for Analyzing Time Series Data Drawn from Complex Social Systems. *PhD Thesis, University of Maryland*, *Supervised by Michelle Girvan and William Rand*. https://doi.org/10.13016/M2V93N

James, R. G., Ellison, C. J., & Crutchfield, J. P. (2018). dit: A Python package for discrete information theory. *Journal of Open Source Software*, *3*(25), 738.

Marton, K., & Shields, P. C. (1994). Entropy and the Consistent Estimation of Joint Distributions. *The Annals of Probability*, *22*(2), 960–977. https://doi.org/10.1214/aop/1176988736