

à paraître dans E. Brousseau, M. Marzouki, C. Méadel (Ed.), 2008, *Governance, Regulations and Powers on the Internet*, – Cambridge University Press

Information communities and open governance: boundaries, statuses and conflicts

Information and Communication Technologies (ICT) produce new meeting points, *on-line communities*, which organize the consumption of information and cultural products by proposing a choice between a large number of proliferating goods, opinions, critiques, news and commentaries. The upsurge of information communities on the Internet is thus linked to the development of a new mode of information retrieval. Unlike instrumental use of a search engine, which consists in a pre-formatted inquiry, information communities and more generally cultural Internet – Web 2.0, participative platforms, folksonomy – support disorderly and inquisitive exploration (Auray 2007). On-line communities play a key role here: apart from trimming and altering huge corpuses of knowledge, they stimulate a more random discovery of the corpus through mechanisms of suggestions and recommendations, or through the exchange of news, advices and commentaries.

In the collective configurations that will be studied here, those online corpus are fed by aggregation of many isolated contributors who come from various backgrounds, and which commitments to the group are often weak. This raises the issue of power and organization in those collectives : do ICT transform, and if so under what conditions, the conduct of governance in the communities? In other words, how do technology perform the organization of those places ?

The type of the communities here studied is based on a the huge size of the involved collectives, the principle of openness, the collaborative production with explicit rules and detailed manuals. The rules are embedded in technical disciplines which can be complex and sophisticated : Wikipedia, ebay, Debian, Slashdot,..., are marked by a chiasm or a structural tension. Firstly, they are based on norms and their *interiorization* by members : sometimes they impose sanctions and penalties but mostly they attribute “prestige indicators” (positive sanctions), based on peer evaluation to their members depending on the quality of their past contributions : “karma” in Slashdot, “stars” in ebay, “hall of fame” in big forums.... However, those collective are not only governed by the interiorization of norms : they are

based on the auto-organization of the collective knowledge production and on *emergence* mechanisms, they claim permissiveness, they try not to force the contributors and to let them free, and they partly function without collective intentionality.

The paper thus explores the conventions that allow the “arrangement” of the collective production ; these conventions are explicated in rules and sometimes delegated to automated devices, with a reliability that can be limited or uncertain. Some rules are more or less collectively based on the principles of democratic governance. The chapter analysis therefore procedural mechanisms of governance of this type of electronic collective. The first part focuses on the demarcation of borders: how to protect these areas in the absence of barriers to entry and even though they suffer from the onslaught of vandals, spammers, manipulators, cheats, voluble committers.... The answer is sought in two directions: the fight against underhand vandalism (1.1) and automation of cleaning (1.2). The second part studies the status of individuals and looks at the devices invented to fix the statutes of members. They may ensure that the status reflects the individual level of contribution of each member, while maintaining the collective nature of the activity (2.1) and they have to secure the reputations (2.2). The third part is about the pacification of conflicts by proceduralization of regulation. One of the issues is the sophistication of the veto (3.1), but the struggle against fighting is focused on the creation of instances of moderation and mediation committees, rather than the deployment of an arsenal of sanctions (3.2).

1. The governance of boundaries

Since, in cultural Internet, Internauts want to browse and thereby serendipitously discover content that they would not deliberately have sought¹, most large on-line collectives are characterized by the fact that they host opportunistic and occasional contributions. They are open to "weak cooperation" (Cardon 2006). For instance, user lists open a frame of intelligibility which makes the most unexpected or surprising statements relevant. They are open to writing, even to non-subscribers. This strange openness is valued for its epistemic benefits: it makes it possible to construct conditions favourable to discovery. Initiatives taken to restrict this openness are subject to acceptance by a qualified majority. For instance, the

¹ On the centring of on-line communities on serendipity and the exploratory paradigm, cf. Auray (2007).

fragmentation of lists is a strictly controlled process. On Usenet, the historical model of discussion forums², a break-away forum has "to have over 80 more YESSSES than NOS and over three times more YESSSES than NOS" to obtain approval. Messages are rejected only rarely, according to a procedure adopted by consensus. Usenet uses the "Breibart index" which calculates the sum of all forums on which the message has been published many times and the square roots of forums on which it has been published, to calculate the threshold over which a user message is banned from a list. This threshold is set by consensus between the members: on the Big8 (the 8 main moderate scales) a robot rejects a message if this index exceeds 20 over the past 45 days. The threshold is reduced to 10 in 30 days in the French scale.

Moderation is carefully watched, as the episode of the influential list of debates between "nettime" activists, which established jurisprudence in this respect, attests. In 1997, two years after its creation, following the proliferation of spamming and flaming, and the deterioration of the signal/noise relationship at a level threatening the viability of the list, moderation was introduced. It nevertheless triggered a conspiracy against power grabbing by moderators. The debate was particularly sensitive around the content of ASCII art and spam-art: should they be considered as destructive flooding or as a contribution in the discursive space? The decision was taken to create an additional list, nettime-bold, receiving all the messages that had not yet been moderated. This made it possible to publicly evaluate moderation criteria, and enabled those who so wished to implement their own moderation methods.

Whereas it may seem more effective for the user who asks a question to have to pay a fee, and thus to relieve the advisers, for instance by adding keywords, it is rarely practised³. On the contrary, regular advisers write FAQs to facilitate new readers' integration. None of the fifty large free software advice exchange projects require the requester to fill in a form or to read

² Usenet by its hierarchical organization and the cultural weight of a kernel of the system administrators in its governance (the « usenet dinosaurs ») is slightly different from the other online collectives that we study in this chapter; but it was a model for most online collectives because it was the first newsgroup network based on procedural rules and governance mechanisms.

³ In their statistical study Lakhani and von Hippel (2000) note that with 400 questions posted per month between 1999 and 2000, and 1,900 contributors, users of the Apache software help forum were satisfied, no question was turned down, and the cognitive charge was left to the information providers. The same applies to Linux User studied statistically in 2002 (Lanzara and Morner 2003), with 2,500 questions per month and over 1,000 active participants, and Debian User French studied statistically from 1997 to 2007 (Dorat, Conein, Latapy & Auray 2006), with 700 discussion threads per month and 4,000 active contributors for that period. Lakhani and von Hippel (2000) note that it was with an increase of a factor of 100 only that the Apache design required an amendment.

the archives before asking a question. Newbies should not have to go through all the archives for an answer before posting a question.

1.1. Automatic quality control

When the corpus is read by a large outside public, as in the case of the on-line encyclopaedia Wikipedia, the rule is to authorize writing access to anonymous contributors. The aim is to encourage participation by gratifying newcomers who see that their corrections are immediately taken into account. Despite risks of vandalism, on the francophone Wikipedia any contributor, even one who is not registered, can write a contribution and even create an article. Only the two largest Wikipedias have started to introduce prior editorial control. Since December 2005 the English-language Wikipedia has been marked by the Seigenthaler affair: a victim of defamation mediatized his complaint of having been defamed in an article where an anonymous contributor presented him as directly implicated in the assassination of Robert Kennedy and his brother, although the author added that nothing had been proved. Even though this kind of vandalism had already taken place, in this case it remained on line for a very long time without any contributor pointing out this defamatory addition⁴. Since December 2005, on the English-language version of Wikipedia, only registered contributors can now create new articles. Registration is nevertheless limited to the creation of a user name, which can be a pseudonym, and a password. Mention of an email address is recommended (if only to recover one's password if one has forgotten it) but not compulsory. On the other hand, there is no need to register in order to correct or complete articles that are already on line. Since September 2007 the German Wikipedia makes instantly visible only corrections made by "trustworthy" users. To gain this status of trust, users have to show that they qualify by producing at least 30 corrections in 30 days. Inexperienced users have to wait for validation by a trustworthy editor for their corrections to be taken into account.

Editorial control is limited to *a posteriori* elimination of reprehensible content, and to mutual surveillance. One characteristic of the governance of boundaries characterized by the choice

⁴ Considering that he had been insulted, and refusing legal action due to US law on the protection of private life, which would make it difficult to ascertain the identity of clients of Internet access providers, he preferred writing on *USA Today* that "Wikipedia is a flawed and irresponsible research tool" (29 November 2005).

of openness is the presence of automatic tools facilitating the tracking of amendments. Robots used by voluntary correctors verify the presence of offensive words in recent amendments. When they spot an amendment made by an IP, they check the “diff” and compare it to a list of expressions, each of which has a score. If the total exceeds a certain threshold, the robot removes the amendment. Tools scan the origin of the IP addresses used to correct an entry, to identify whether these corrections come from large organizations which may want to "fiddle" information to suit themselves. A major governance problem is control of the consequences of the activity of these automated robots. For example, a central rule has been decreed, the "1RR" (1-revert rule) to "train" the Wikipedia robots so that they do not remove the same IP on the same article twice. The bot can ignore this rule if a human also removes the IP, which creates a bot, human, bot, human, etc." cycle. These limits embedded in the robots' coding are designed to give the final say to the people using them.

Controlling automatic quality control tools

Controlling quality *a posteriori* is based on tools used to retrieve *more stable* versions from on-line corpuses. Thus, the production and freezing of "stable versions" from a selection of content is based on automatic tools like annotating robots in Wikipedia, which draw up a list of eligible articles for a stabilized version⁵. In Debian, for example, a software code is automatically transferred from the *unstable* development version to a stable *test* version, if it has been in the archives for over two weeks and if a critically serious bug causing deadlock has not been attributed to it during that period. This process can lead to considerable perverse effects by giving a "bonus" to the unpopular packets. The packets that rise fastest to the "stable" status correspond not to the best ones, but to those that are the least often looked at. As they are not subjected to the sagacity of a large number of "proof readers", they slip through the net. This problem of the quality cycle is crucial, for quality is precisely the main theme of open source software.

Thus, in the *design* of these automatic editorial control tools human control is unavoidable. On Wikipedia, for example, the criteria to cover for labelling these weak corpuses are a subject of debate. It has to be established which algorithm should be used to mix the progress criterion, measured in terms of the robustness of the content with regard to recent

⁵ Wikipedia in German, the first to have taken the initiative, produced its own DVD-ROM of over 620,000 articles in November 2007.

amendments, and the criterion of the importance of data, measured by the number of visits to a page, for example. On Debian, because of the limits of automatic tools, a qualified majority voted for a human procedure to decide on the "stable" version: the "freeze" phase. The members of the collective meet at a time appointed by the project leader, for an intense cycle of tests leading to the "freeze" of the distribution. This cycle is organized as a series of fun, social events called "Bug Squashing Parties" on the IRC charts (Auray 2004), where a large number of developers synchronize. The choice of elements that will make up the stable version is thus the result of a collective decision.

The impossibility of completely automating the correction work is one of the main bottlenecks. Since 2002 Debian participants have considered the long interval between two stable releases, and the impossibility of controlling the constant increase in the time taken for correction during the test cycle, to be the main problems facing the community. This impression of having reached a critical mass in the project has triggered procedural innovation in the past few years, including the creation of a voluntary quality-control group to deal with these issues, and the proliferation of suggestions to restrict the editorial content. But notwithstanding these proposals, the problem has persisted. The difficulty can be solved by entrusting it to organizations with paid employees, supported by the on-line community.

1.2. A struggle against underhand vandalism

Unlike collectives based on an existing community of values and intentions, large on-line collectives are based on weak cooperation characterized by the formation of "opportunistic" links which do not imply collective intentions or a sense of belonging to a "community". The contribution is often casual, largely unintentional, fragile and temporary, and the low level of mutual acquaintance generates risks of vandalism. One type of vandalism that is easy to curb is the provocative kind. It is based on a logic of pranks, for instance when an Internaut replaces the photo of the pope in Wikipedia by that of Dark Vador. The trickery is temporary and intended more as a joke than to manipulate. On the other hand, the fact that few contributors are acquainted generates a more harmful kind of vandalism, one that is underhand. This kind is based on the manipulation of information which is intended to *remain invisible*. Thus, on-line collectives are often infiltrated by real organizations which manipulate general opinion to run down a rival, or to promote their own products or themselves on the

Internet. For example, a 25-year-old neurobiology student, Virgil Griffith⁶, developed "informers" to draw up a systematic list of the main manipulations: discreet falsification of words, change of adjectives, etc. The sources included private firms as well as public authorities (municipalities, states), political organizations and sects.

Underhand vandalism is an intentional falsification of content, which is supposed to be taken seriously and to dupe readers. Yet the line is fine between strategic manipulation and the contribution of content that enhances general knowledge. In most on-line communities a writing principle is the openness to the members' personal points of view. This is related to the fact that everyone expresses the excellence differently. Even Wikipedia, notwithstanding its encyclopaedic principle of "neutrality", defines objectivity in relative terms, as "the sum of relevant points of view on a subject". Contributors are encouraged to introduce a subjective point of view, provided that it is linked to its enunciator and presented impartially. The epistemological grounds for this approach are a subjectivist conception of truth: stabilized scientific facts are presented as statements related to a point of view, whereas controversial theses are cited as relevant to a subject. *"What people believe, that's the objective fact"* (Wales, 1998). As it is legitimate to add to the corpus by representing a personal point of view, underhand vandalism is a difficult enemy to combat.

Has it moreover really been proved that combating strategic manipulation will enhance the quality of forums? Some models of discussion forums on cultural goods (Dellarocas 2004) show that an eradication of strategic manipulations could have a globally negative impact on the site by removing incentives to contribute and thus decreasing the amount of information produced.

The governance of on-line communities is based on two objective criteria to isolate underhand vandalism from controlled partiality which, on the contrary, is welcome. Thus, *unilaterally deleting* contributions with which one disagrees is considered as vandalism and therefore prohibited. Likewise, the repeated deletion of content (like the Church of Scientology which deletes paragraphs criticizing it on Wikipedia) is banned as a "breach of publication" on Wikipedia and is punishable by banishment from the account. The systematic reiteration of a contribution is also considered as underhand vandalism. Other cases are not controlled.

⁶ <http://virgil.gr/>

The stakes are high, for underhand vandalism pollutes the corpus and can cause readers to lose interest. In fact it can result in an insidious cleansing of the on-line community by an activist minority representing a particular point of view. The natural form of underhand vandalism is thus strategic infiltration. Many on-line collectives have had to deal with it. For example, on Usenet it was necessary to combat manipulation of discussion forums by sects. A decisive form of control was the setting of rules to define who is eligible to vote on a subject. Since voters' identity is difficult to establish in on-line collectives, there is a risk of infiltration. For instance, a founding principle of Usenet was that, to vote on the creation of a new discussion forum, anyone interested in the topic could vote, which included people who were *not yet users*. This led to infiltration during the fr.soc.sectes affair on Usenet, between September 1997 and March 1998. The vote for the creation of a discussion topic on sects was held in September 1997 and was marked by the jamming of the ballot boxes. There was massive and organized dissemination of a call to vote No, from members of the Moon sect. Mass mailing to sympathizers of the sect was intended to encourage them to vote "no", on the basis of information that was biased to say the least, and in any case without any relation whatsoever with the original charter of the fr.soc.sectes forum. Some answers to the acknowledgement of receipt of the ballot papers indicated that some of these people did not speak French and therefore that they were unable to understand the RFV⁷. The trickery was considered as evident and likely to have changed the result of the vote. Since then, Usenet has strictly limited the right to vote on the creation of forums. Only informed participants, who fill in a form and thus show that they speak and understand the language and the Usenet rules, may now vote.

The principle of openness of the on-line community's borders is therefore strictly coordinated. This consists in *a posteriori* cleaning of the content, resembling a system of editorialization (producing stable versions with content selected by automats); and a moderate control of strategic manipulations, on the basis of a war on underhand vandalism and hostage-taking by activist minorities.

⁷ *Request For Voting*. On-line collectives also have to fight against strategic misuse of voting systems. For instance, during the elections for a project leader, members of Debian had to protect themselves against misappropriation of the "scorings à la Condorcet/Borda" (votes where all the candidates are arranged by order of preference). Some votes, for example, classified last the candidate who threatened their favourite. Thus, the wish to struggle against insincere votes is the justification for the change of variant of the Condorcet election, with the replacement of the "single transferable vote" by the "Schwartz sequential dropping".

2. The governance of status

On-line collectives propose an original articulation between individualism and solidarity. They favour a common good dynamic based on a logic of personal interest, and change members' statuses in accordance with their internal contributions. They highlight the most prestigious statuses by displaying the names of people who excel on a merit list and granting them exclusive rights. Statuses stem directly from internal contributions. The originality of large on-line self-organized collectives thus relates to the endogenization of statuses around *profiles*. A pseudonym is the person's name; a profile is the evaluation of that pseudonym on the basis of past transactions⁸.

There is nevertheless a difficulty in enabling these individual scores to emerge. How can contributions be individualized in a space, that of on-line communities, based on the collective character of the outputs? An encyclopaedic article in Wikipedia, a discussion or a database are based on a sum of complementary individual inputs, and the quality is based on the aggregation of the number plus the individual performance (Surowiecki 2004). The choice of ranking individuals differs according to the social and relational complexity of the collective knowledge constructed by on-line collectives. Thus, collectives based on diffuse marks of reputation for individuals can be distinguished from those based on objective scores. The former, like discussion forums, elaborate relationally highly complex technical knowledge, in which individual contributions are entangled in a network of expertise in which it is impossible to detach "discussion threads" (Conein 2005). The latter, like video platforms, cluster individual content, thus allowing for individuals' evaluation, expressed in the form of cardinal numbers and coloured icons.

For example, lists and discussion forums are not simply places for exchanging opinions and advice. They are also places where certain experts build up their authority (Welser and Smith 2007) through mechanisms of selection of partners during discussions. The presence of a core of uncontested experts and the average skills of participants on *Debian* lists, for example, exert a strong constraint on the production of questions and the selection of those who answer

⁸ For example, in *JDN(1211)*, the pseudonym is *JDN* and the profile is *1211*. This means that the person has a score of 1211 points for the quality of his or her transactions.

(Conein 2006). Their scores nevertheless remain diffuse. The reason why discussion lists are limited to diffuse marks of status is that it is difficult to credit a particular message in a discussion as a "good answer". The result is generally collaborative, and the cognition is distributed, borne by the entire chain. This is illustrated by the relative failure of the "Yahoo Answers!" site⁹ which functions with a system of points and levels. The number of points obtained increases when a contributor has been chosen as the "best answer" in the discussion (paradoxically, it also increases for someone who has asked a question for which no best answer has been voted). The points make it possible to access levels which attribute rights. As the answers are collaborative, the "best answers" system has led to misuse of the site, which deteriorated into a space for recreation, riddles and jokes. Riddles create competition between participants, who are reduced to choosing the funniest answer, so that there is no longer the collaborative aspect. There is a simply a duel between rivals, for the fun of it.

Other on-line collectives, in which individual contributions are easily individualizable, associate icons and scores with members. Some of these propose "signs" which enhance visibility. Having a large number of stars on eBay or a good karma on Slashdot means that one's visibility or comments move up in the on-line community and give one a decisive advantage as regards reputation. Through this self-amplified mechanism, a bonus of recognition accumulates on those that have already acquired it. The other mechanisms are "ranks" which afford access to different powers. In open source software projects (Free BSD) or Wikipedia, a minimum number of contributions enable one to be eligible for the rank of "officer" or "administrator" which confers rights on the deletion of accounts or content. The promotion is however not always automatic¹⁰, it is subject to an additional vote. In view of the importance of these signs of status for the members, a second level of governance is designed to control the *fairness* of the reputation scale. It focuses in two directions. First, it aims to control the effects of these community members' reflexive knowledge of the calculation algorithms that produce reputation, as this knowledge leads to biases that constantly have to be corrected. Second, this governance aims to control and sanction the effects of reputation on the members' behaviours.

⁹ <http://fr.answers.yahoo.com/>

¹⁰ Although it may be. Since September 2007 the German Wikipedia has given its contributors "indexes of trust". Only the corrections of "trustworthy" contributors are immediately visible. The software indicates the contributions in different colours, according to the author's index of trust, which is based on the author's past contributions: those whose contributions are seldom amended obtain a high index, whereas those whose contributions are amended soon after being posted are given a lower index.

The governing of activity by public evaluation generates problems concerning the *design* of the status scales. The *fairness* of these scales has to be controlled, by avoiding biases and manipulations of their score by individuals (2.1). At the same time it is essential to make identities and profiles secure (2.2). The communities have to be protected from the threat of crooks who, by usurping profiles, can undermine confidence in the profiles which the sites allocate to their members.

2.1. The effects of reflexive knowledge on the calculation of reputation

Because they are so important, marks of reputation can easily have undesirable effects in on-line communities. Members obsessed with the consequences of their behaviour on their virtual aura may alter their habits or refuse discussion. Yamauchi et alii (2001) showed that in large software development projects, interactional patterns are biased towards action¹¹. The tasks performed are not preceded by a declaration of intent. This bias towards action is explained by its protective aspect for reputation. By not stating their personal engagement in a precise task of the "to do list", contributors protect their face. "I can get it wrong because I don't tell the others what I'm doing" (a Debian developer, Sept. 2005). Those who have little confidence in their know-how see barriers to their engagement, which would be very strong if they were obliged to publicly state what they were doing.

The main distortions are however related to the actors' anticipation of the consequences of their evaluations on their own reputation score. The site for sales between private individuals affords a good example of "bias" related to actors' knowledge of the mechanisms of calculation of reputation. Each member of eBay has an evaluation profile based on their interlocutors' evaluations. Buyers and sellers can mutually assess one another by posting an evaluation for each transaction. It is the buyer who starts, by posting his or her evaluation, and

¹¹ Using transition diagrams based on the coding of language acts in messages of the general discussion forum, Yamauchi et al. (2004) show that the probability of transition from "report" to "question" is far greater than that of the inverse transition (of "question" to "report"), which is negligible. (Yamauchi and al. 2001) Discussions take place after the report on an action, not prior to it. The category with the highest frequency in initiating a thread is "report" (32.3%), before "question" (22.6%). Moreover, only 14.4% of the "reports" lead to a "question" (and thus to a discussion), whereas 28.9% have no critical follow-up.

after the seller has received it, he or she posts his or hers. These evaluations cannot be altered or erased (only comments can be deleted, by common accord). Many buyers therefore prefer not to leave a negative or neutral evaluation when they have had a bad experience, for fear of receiving a negative evaluation by the seller. Negative evaluations out of retaliation have thus become common practice. Consequently, a "kind bias" has appeared (Dellarocas, Fan and Wood 2004): the number of negative or neutral evaluations left on the site is smaller than the number of bad experiences signalled¹². Reflexive knowledge of the calculation of the evaluation leads to bias.

The governance of status aims to systematically correct these effects of reflexivity by constantly adjusting the system so that it corresponds to the acts performed. Since May 2008 e-Bay has reformed its evaluation system: sellers may no longer leave negative or neutral evaluations of buyers. Certain sites have sophisticated auto-regulatory mechanisms to constantly ensure the convergence of mutual evaluations towards honesty. For instance, the Slashdot news site is based on mutual control of evaluations by members, by means of meta-moderation. If a member has their moderation noted as "unfair", then the number of points that they can use for their power of moderation immediately decreases. The system is auto-regulated, in so far as the meta-moderators are recruited among the moderators with the highest score.

2.2. Security of profiles

Sites can turn into hell if pseudonyms are stolen or are publicly slandered without justification. For instance, some celebrities have discovered a "virtual clone" on MySpace, steered by someone else and based on a mixture of true and false, of information taken from

¹² As a user of the site notes, "In mid-August I received an email from a seller who didn't follow up. I tried to negotiate but it wasn't possible so I filed a complaint on eBay. I logically posted a negative note on the seller. In retaliation, I assume, the buyer posted a negative note on me! That beats it all! My question is therefore: if a buyer/seller who is a victim of a dishonest buyer/seller can't take the risk of posting a negative note, in case a negative one is posted on them in return, what is evaluation worth?" (forum on the e-Bay evaluation system, 5/3/2007).

the Net and personal pseudo-details that are mostly false. Where does the common knowledge end and the usurpation of identity start?¹³

Apart from this threat of usurpation concerning famous people, any individual with a good reputation in an on-line community is at the mercy of a pseudonym thief attracted by the prestige. This is how "phishing" develops. On e-Bay, the fact of being able to have an opinion before betting or selling is crucial. This explains the decisive importance of each member's profile, including their evaluation history and a score that sums up the trust that can be granted to that person. The score is the product of an evaluation system which enables each member to mark and to comment on the buyers and sellers with whom they have carried out transactions. To clear their reputation certain crooks use a phishing technique: they inundate users with false emails, claiming to be the customer service of e-Bay and asking them to confirm their pseudonym and password because abnormalities have been found in a maintenance operation. Of the thousands of emails sent, sometimes a naïve user, or one who fears being excluded, provides the requested information. The crook then takes over their account and alters the password so that the legitimate owner is dispossessed of it. This false e-Bay email comes from addresses that cannot be traced, with links that do not remain on the servers – which are often unaware of hosting them.

On-line communities are all particularly severe with these usurpers. The sanctions are maximal: "Death Penalty" on Usenet, i.e. life banishment from the IP address; "permanent disqualification" on Wikipedia, i.e. the IP addresses used by the member are banned from contributing. Legal measures are rare. It is nevertheless almost impossible for the different sites to permanently ban a person, as they can always reappear with a new address. Only legal action, which implies that charges are laid and that the access providers cooperate, can definitely stop such crooks. But legal identity is badly protected these days. In France, pretending to be someone else, by "phishing", is not considered as an offence since on Internet usurpation concerns "logical identities" (IP, login, email) and not "real" identities such as surname, first name, address¹⁴. These logical identities – including the usurping of an

¹³ An example is the "false blog" of Jacques Chirac. Likewise, in an article on the misappropriation of a virtual clone of a famous philosopher on the Web, Judith Revel wonders what auto-regulatory bodies can be conceived to avoid people finding themselves with racist or provocative clones (*Libération*, 8 November 2007).

¹⁴ Senator Michel Dreyfus-Schmidt has tabled a bill in parliament on the subject but it has not yet been passed.

email or IP address – are not protected as such. Likewise, the possibility of protecting the profile by virtue of a "property right" is not yet established in practice. To offset this shortcoming, sites use indirect means, invoking fraud through alteration of the truth likely to cause damage, or public defamation.

3. The governance of conflict

On-line communities are characterized by regulation which aims to deal with the maximum number of problems closest to the ground. Like the struggle against vandalism, regulation of large cooperative collectives is based on mechanisms of mutual surveillance. Whereas in debates on on-line cooperation the focus is on participative writing, it is *participative surveillance* that seems specific and new in community sites producing knowledge. Most discussions are maintained at the lowest level. For instance, in discussions on Wikipedia, anyone has the possibility of criticizing others¹⁵. It is only when mutual surveillance no longer suffices that additional regulatory procedures are introduced: mediation and sanctions. Under the effect of a "warning", mediators are brought into action. A *public* space is then constituted, where the problematical article is shown to a *third party* that is not a stakeholder in the conflict. The problem is then discussed on a special page examined by voluntary *watchers*. If the authors diverge again, in residual cases, a third regulatory body is called on: the arbitration committee, consisting of members elected by peers, to which any contributor may refer a complaint, and which is committed to considering it as admissible and to proposing a sanction.

Conflict governance aims to maximize the possibility of this direct participation and the production of consensus, without it undermining or disrupting the activity of contribution (3). It functions in two ways. First, the idea is to avoid *paralysis*: how do these advanced communities deal with the highly time-consuming development of a large number of calls to discuss or vote on? How do they cope with the risks of deadlock related to the fact that consensus has to be reached? (3.1). The second aim is to avoid the *degeneration* of discussions into fights. It is necessary to counter the constant risk of public argument and

¹⁵ The weight of this regulation is relatively weak: on the anglophone Wikipedia, *only 14% of pages have a discussion* (Levrel and Poudat, 2007).

controversy deteriorating into *ad hominem* conflict characterized by anger, humiliation and an incapacity of the different partners to respect the standards of public argumentation (3.2).

3.1. The struggle against paralysis

The place granted to participation is subject to the threat of *paralysis*, related to the cost in time of having to put issues to the vote. Due to the asynchronous nature of interactions on the discussion list that frames the meeting, the duration of the vote lasts for at least several weeks – which is long¹⁶. A constant concern of sovereignist on-line communities is, paradoxically, to ensure that they are not submerged by too many applications to vote on¹⁷. How do these advanced communities avoid the over-consumption of time? First, they try, as far as possible, to limit the number of votes to a strict minimum. The *vote* must therefore be held only after a discussion has taken place between all the interested participants, over a sufficiently long period (the minimum interval before a call for a vote can be made is for example two weeks, on Debian). This makes it possible to clarify the debate, to delete the craziest options, and to reach consensus. In general, on the most controversial points the discussion is accompanied by "mini-polls" which do not have the value of a final vote, but are intended to test the balance of powers. They do not make the decision, they inform members. Virtual participative communities also implement another technique to save time. They submit the votes proposed to a *quorum* (for instance in Debian the quorum equals 1.5 times the square root of the number of developers, so that with 800 developers, the quorum is 43 voters). But this quorum is stricter than in the traditional sense. The community members vote by classifying all the options by order of preference, according to the Borda method: all the options are compared in pairs, and the blank vote, "none of the above",¹⁸ is included in the options. The quorum is considered as having been reached when the choice proposed is classified before the choice "none of the above", in a number of votes greater than the quorum. Thus, this special quorum not only controls the fact that the *mobilization* of the electoral body is sufficient; it also

¹⁶ The duration of the vote on the French ranking of Usenet has a minimum of 21 days and a maximum of 31 days. On Debian it is two weeks.

¹⁷ As a Debian developer explained: "We want to have as few votes as possible to settle an issue, since each vote requires two weeks to run in order to get the most input. This means that we can't follow a traditional procedural amendment process – each vote has to have all proposed (and seconded) amendments on it, and the procedure has to select from among them" (Buddha Buck, 11 June 2003, Debian-vote list, summary of his proposed amendment).

¹⁸ In the discussions this option is called "*further discussion*" or "*keep talking*".

controls the *determination* of the voters, above a certain threshold. In the classical conception of the quorum, abstention is taken into account; in this sophisticated conception, the voters' determination is taken into account.

Finally, virtual communities are vulnerable to deadlock because individuals have the *right to veto*. For this purpose flexible rules have been improved. These systems are sometimes called systems with a *relative* right to veto, as in the case of Wikipedia¹⁹ where the veto may be waived in cases of a very strong majority, thus limiting the risks of deadlock. The procedure consists of a first vote during which participants can be for, against or against with a veto. After the vote, the number of "for" is counted and compared to the total. If there is no veto, the qualified majority validates the vote (e.g. 70%). If there is a veto, a second discussion may follow (depending on the subject) to try to identify other options. The discussion is then followed by a second vote with a higher qualified majority than for the first vote (e.g. 80%). If the second vote ends with a veto and 75% for, the against prevails. If the second vote ends with a veto and 85% for, the for prevails. With this system the veto can be taken into account, but its waivering is nevertheless authorized in case of a very strong majority, thus limiting risks of deadlock.

It is interesting that this system was imported from real and longstanding democratic practices. The participant who set it up for Wikipedia, Aurevilly, noted that: "it's the most common political system when a right to veto is granted to an individual on the basis of collegial deliberation, in order to avoid a single individual definitively blocking the expression of the general will. We sometimes talk of a right to new deliberation instead of the right to veto: the veto can be waived by a new unanimous deliberation of the assembly. This is more or less the system in the US, where the President can veto a law passed by Congress, which can in turn bypass the veto by voting its cancellation with a two-thirds majority. There is also a veto which is so-called relative in terms of duration, such as the one granted to Louis XVI: the veto suspended the deliberation voted for both legislatures" (French Wikipedia, 9 April 2004). This is a fine attempt to differentiate a veto from a strong opposition.

On-line communities seem to be evolving towards participative mechanisms, but these are characterized by prevention against paralysis due to an excess of votes and vetoes. Their

¹⁹ The Debian notion of super majority follows the same pattern.

reflection on the place and function of the vote make them rich lessons for understanding the current mechanisms of consultation, which often shed little light on the moment at which the collective decision is actually *taken*.

3.2. The struggle against fighting

Over a certain threshold, discussions in on-line communities tend to become illegible and the risk of deterioration increases. There is a saying for this, inspired by Mike Godwin, a pioneer in on-line communities. Godwin's law is an extension of the *Reductio ad Hitlerum*, which is part of the Usenet folklore. In 1990 Mike Godwin declared the following empirical rule: "As a Usenet discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches 1". The longer a discussion, the greater the chances are of just anything being mentioned. Moreover, a point is reached where the conversation irreversibly sinks into cacophony and fighting, and the probability of reverting to the initial subject in an impartial tone approaches zero.

This statistical finding can be explained. There are civil forms of expression that have to be used in any electronic discussion: one has to remain impartial, respect one's opponent, maintain a convivial tone, and always try to avoid misunderstanding. Electronic conversation, which shortens the interval between the intention and the act (messages can be sent at a click), and leaves lasting traces of all dialogue, makes it difficult to maintain these rules of courtesy in the dialogue and debate.

As regards governance, the methods employed by communities to avoid deterioration consists in favouring moderation and *mediation* rather than sanctions. As soon as a debate reaches stalemate, a third party who is not a stakeholder in the conflict is automatically brought into the picture and can thus act as an impartial spectator or even an arbitrator. The idea is to switch from a confined space to a public one, and thus to urge the protagonists to improve their behaviour because they are being watched by a moderator who may firmly ask them to stop arguing. On Wikipedia participants in a discussion that becomes bogged down are invited – by means of a warning on a banner on the page – to go to a public page where the discussion will be examined by *watchers*. This switch to a public space encourages moderation in discussions. It is accompanied by a reminder, through links, of the rules

(policies) which encourage self-discipline. Yet the governance of communities cannot simply decree "policies" framing individual activities, and leave it up to members to take the initiative of switching to mediation as the need arises. Two governance mechanisms are essential to limit fighting: the supervision of discussions as soon as precise warning signals appear; and the establishment of referee committees to decide between the contradictory interpretations of the policies.

On-line communities cannot simply display a warning banner and leave it up to their members to take the initiative of bringing a controversy into the open. The communities identify objective signals, easily detectable by robots, which warn that a debate is degenerating. The fact that these indicators can easily be traced and thus detected by robots facilitates governance and allows for partial automation. On Wikipedia, for example, which is on the cutting edge in this respect, the crucial sign that will trigger the switching of the debate onto a public page is the existence of a cycle of mutual deletion of contributions. The so-called "3 reverts" rule requires that a contributor may not delete contributions of another editor in an article more than three times during a period of 24 consecutive hours. The size of the amendment deleted is irrelevant. Any violation of this rule triggers a warning banner and the vigilance of a squad of volunteers who control the dispute. The community sees this rule as an "electric fence", designed to defuse publishing wars as soon as they emerge.

Furthermore, even when the discussion is placed under the control of public authorities, just the presence of policies is not enough to moderate the dispute. One can even consider that, on the contrary, the wide diversity of policies – 42 policies and 24 guidelines for the Anglophone Wikipedia – is a source of dissent, because each of these rules is ambiguous and can therefore be interpreted differently in the context. Moreover, the invoking of contradictory policies encourages power plays between the actors (Kriplean and Beschastnikh et alii. 2007), which is why arbitration committees are set up. For example, on Wikipedia this committee undertakes to receive complaints from contributors who are unable to agree on the interpretation of policies. The referees have to undertake to propose a sanction and to agree between themselves on the adoption of sanctions. If unanimity is not obtained from the outset, they vote three more times, and if it is still not reached with the fourth vote, they accept a sanction agreed by the majority in the following vote.

The governance of conflict is thus based on a model of three-tiered regulation, one of the most complete versions of which is on Wikipedia. The most formal conflict management mechanism is thus based on "bottom-up" management of most conflicts, by virtue of procedural rules. When this conciliatory solution proves to be inconclusive, the disputes are referred via partially automated procedures to mediators. And when these prove to be incapable of solving the disputes, because of character problems or acrimony, an arbitration committee can decree dissuasive sanctions. The other large on-line collectives function much in the same way, even though they are not nearly as formalized. Thus, the communities producing open source software are based on discussions which seek consensus and, when opinions carry on diverging, they bring in the mediation of technical committees: the Technical Resolution Board (9 people elected by accredited developers) in Free BSD, and the Technical Committee in Debian (8 members). It is only when an agreement is still not reached that the subjects are submitted for plenary votes, or proposals to exclude members are formalized and submitted to the decision of the appropriate authorities.

Conclusion

In the information communities taken as examples in this article, many of which attract a huge number of contributors, the regulatory mechanisms aim to fight not against an under-supply but against an *excess* of participation. The governance is centred on cleaning out the corpus, chasing vandals and organizational submarines which want to surf on the site's popularity to defend their own interests. The idea is also to combat redundancy and natural tendencies, when the corpus increases, towards illegibility, deterioration of quality, and the decline into a generalized fight involving many controversies.

The governance methods thus have to respect the principle of openness to outsiders of information communities. They are therefore oriented in three directions. First, the governance of borders is based on *a posteriori* quality control of *stable* versions, by systems of selection, within the corpus, and without affecting the way in which it is fed. This *ex post* selection, based on partial automation as well as on the struggle against underhand vandalism and infiltration by activist minorities, focuses on seeking out surreptitious deletions from the

corpus. Second, the governance of status is based on the granting of a high level of symbolic recognition to individual participants. There is an adjustment of "formats" of this reputation (Thévenot 2007) to the type of corpus produced by the on-line collective: the less individualizable the contributions are, the more their marking is diffuse. The governance of status seeks to dynamically correct the biases introduced by the members' reflexive knowledge of the mechanisms for calculating the status score. Third, conflict governance is based on the preference for participative surveillance involving grassroots contributors as much as possible. This participative requirement causes the mechanisms of conflict avoidance to be centred on the struggle against paralysis, involving original mechanisms of quorum by option and relative veto, and on the struggle against fights, involving automated mechanisms of switching towards mediators when conflicts turn sour.

Information communities are thus the emblem of new forms of organization which are highly attentive to individual singularities: they articulate group projects to individual status. Today they find themselves confronted by a growing challenge to their governance methods. Whereas their success has been related to the fact of attributing symbolic recognition to their members (prestige, renown), will this be enough tomorrow? Will it not be necessary to *remunerate* certain contributors? And through which mechanisms? Today, with certain contributors' denunciation of the fact that platforms exploit the content of their voluntary contributions, offers of partial remuneration of contributors by communities are proliferating – with Deezer, Bebo.com, etc. What is being lost and what is being gained with this trend? While it may seem legitimate to remunerate the authors, will this not affect the prevailing modes of governance? Is it necessary to generalize remuneration to cleaning and corrective activities, which are of little value to the ego? Is it legitimate to remunerate contributors who advertise by way of their contributions to information communities? Should that which made the success of these communities not be maintained on the fringes: the toys of glory?

Bibliographie :

- Adler, B., Chatterjee, K., Alfaro, L., alii, "Assigning Trust to Wikipedia Content", 2008, *Technical Report UCS-SOE-08-07*, School of Engineering, University of California, Santa Cruz, CA, USA.
- Alfaro, L., 2008, "Measuring Author Contributions in Wikipedia", WikiSym 2008, *International Symposium on Wikis*.

- Adler, B., Chatterjee, K., Alfaro, L., alii, "Assigning Trust to Wikipedia Content", 2008, Technical Report UCS-SOE-08-07, School of Engineering, University of California, Santa Cruz, CA, USA.
- Alfaro, L., 2008, "Measuring Author Contributions in Wikipedia", WikiSym 2008, *International Symposium on Wikis*.
- Auray, N., 2004. "Le sens du juste dans un noyau d'experts. Debian et le *puritanisme civique*". In Proulx, Massit-Follea & Conein. *Internet : une utopie limitée*, Presses Universitaires de Laval, Canada.
- Auray, N., 2007. "Le modèle souverainiste des communautés en ligne : impératif participatif et désacralisation du vote". *Hermès*, 47, pp.137-45.
- Auray, N., 2007. "Folksonomy : a New Way to Serendipity". *Communications and Strategies*, 65, pp. 67-91.
- Avery, C., Resnick, P. & Zeckhauser, R., 1999. "The Market for Evaluations". *American Economic Review*, 89(3) pp. 564-84.
- Boyd, D., 2002. "Faceted Id/entity: Managing Representation in a Digital World." Cambridge, MA: MIT Master's Thesis. August 9.
- Cardon, D., 2007, « 10 propriétés de la force des coopérations faibles », *InternetActu*, 8 février 2008.
- Cassell, J. and Ryokai, K. (2001). "Making Space for Voice: Technologies to Support Children's Fantasy and Storytelling." *Personal Technologies* 5(3): 203-224 Conein, 2006,
- Conein, B., 2005, "Relations de conseils et expertise collective: comment les experts choisissent leurs destinataires sur les listes de discussion", *Recherches sociologiques*, 3.
- Conein, B., Latapy, M., 2008, "Usages épistémiques des réseaux électroniques", *Sociologie du travail*, Vol 50 - N°3.
- Davis, J. P., Farnham, S. D., Jensen, C., 2002. "Decreasing online "bad" behavior", Conference on Human Factors in Computing Systems, CHI '02 extended abstracts on Human factors in computing systems, Minneapolis, Minnesota, pp. 718-719.
- Dellarocas Fan Wood, 2004, "Self-interest, reciprocity and participation in online reputation systems". MIT Sloan working paper, 4500-04, <http://ssrn.com/abstract=585402>.
- Dellarocas, 2006, "Strategic Manipulations of Internet opinion Forums: Implications for Consumers and Firms." *Management Science*. 52(10), pp. 1577-93.
- Donath, J., 2001, "Mediated Faces", in Beynon, M., Nehaniv, C.L., Dautenhahn, K., (éds), *Cognitive Technology : Instruments of Mind*, Proceedings of the 4th International Conference.
- Dorat, R., Conein, B., Latapy, M., Auray, N., 2007. "Multi-level analysis of an interaction network between individuals in a mailing-list", *Annals of Telecommunications*, 62(3-4).
- Fourmentraux, J.P., 2008. "L'œuvre en devenir. Sociologie de la communication et des échanges créatifs à l'heure du numérique", *Communication*, 26(1), pp.108-26.
- Gensollen, M., 2006. "La culture entre économie et écologie : l'exemple des communautés en ligne". In Greffe, X., ed. *Création et diversité au miroir des industries culturelles*, La Documentation Française, Paris.
- Ghosh, R.A., Glott, R., Kreiger, B. and Robles, G., 2002. "The Free/Libre and F/OSS Software Developers Survey and Study—FLOSS Final Report".
- Hardin, G. 1998. "Extensions of "The Tragedy of the Commons"". *Science* 280 (5364): 682–683.
- Hirschman, A.O., 1970. *Exit, voice and loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.

- Kollock, P., 1997. "Design principles for Online communities", in *The Internet and Society: Harvard Conference Proceedings*. Cambridge, MA: O'Reilly & Associates.
- Lakhani, K., von Hippel, E., 2000. "How Open Source software works: 'free' user- to-user assistance". MIT Sloan School of Management WP #4117.
- Lampe, C., Johnston, E. and Resnick, P., 2007. "Follow the Reader: Filtering Comments on Slashdot". Proceedings of CHI 07 Conference on Human Factors in Computing Systems. San Jose, CA, pp. 1253-62.
- Lampe, C., Resnick, P., 2004. "Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space". In Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems, Vienna Austria.
- Lanzara, G.F., Morner, M., 2003. The knowledge ecology of open-source software projects. Proceedings of the 19th EGOS Colloquium, Copenhagen, July 3–5
- Lerner et Tirole, 2002, "Some Simple Economics of Open Source", *Journal of Industrial Economics*, Vol. 50, No. 2, pp. 197-234, June 2002
- Mathes, A., 2004. "Folksonomy – Cooperative Classification and Communication through shared Metadata", working paper, *Computer Mediated Communication*, Graduate School of Library and Information Science, University of Illinois Urbana – Champaign.
- Rehn, A., 2004. "The politics of contraband. The honor economies of the warez scene", *The journal of socio-economics*, 33, pp.359-74.
- Stalder, F., Hirsh, J., 2002. Open Source Intelligence. First Monday, 7(6). http://www.firstmonday.org/issues/issue7_6/stalder/index.html.
- Thévenot, L., 2007. "Reconnaissances : avec Paul Ricœur et Axel Honneth." In Delacroix, C., Dosse, F., Garcia P. (dir.). *Paul Ricœur et les sciences humaines*. Paris, La Découverte.
- Turkle, S., 1995. *Life on the Screen: Identity in the Age of the Internet*, Simon and Schuster.
- Yamauchi, Y., Yokozawa, M., Shinohara, T., Ishida, T., 2000. "Collaboration with Lean Media : How Open-Source Software Succeeds", in *ACM Conference on Computer Supported Cooperative Work*, Philadelphia, December.
- Wales J., 1998, "Neutrality of point of view", contribution sur l'article de la branche anglophone de Wikipedia, <http://web.archive.org/web/20010416035757/http://www.wikipedia.com/wiki/NeutralPointOfView>
- Welser, H., Gleave, E., Fisher, D., Smith, M., 2007. "Visualizing the Signatures of Social Roles in Online Discussion Groups", *Journal of Social Structure*, vol.8.