

Gloria Origgi  
CNRS – Institut Nicod

*Designing wisdom through the Web: The passion of ranking<sup>1</sup>*

*Civilization rests on the fact that we all benefit from knowledge we do not possess*

Friedrich Hayek

With the outstanding potential of the Web 2.0. of aggregating socially distributed information to obtain intelligent outcomes, the very idea of *collective intelligence* has entered a new era. What was not much more than a curiosity in the history of thought, that is, that the aggregation of many unsophisticated judgements may sometimes result in a more sophisticated and precise result than the one produced by a single expert mind, is nowadays a major *fact* that is revolutionizing our everyday way of thinking and taking decisions and needs to be taken seriously. *Google* searches, *Wikipedia* entries, *eBay* transactions, *CiteUlike* libraries, are all based on the simple correlation between the data filtered by the many and the relevance for us of the information we get. Yet, a major concern with these new collectively intelligent systems is that our control over the path of aggregation of information is sometimes poor, and the individual or institutional capacities of intervention on the design of the aggregation process are often very limited. I think we should take the *design* issue very seriously in order to be aware of the biases of these systems and their potential misuses: each system of aggregation of private information and preferences has its special design that has impact on its outcome. And this is a general point about institutional design. No matter how many people are involved in the production of a collective outcome – a decision, an action, a cognitive achievement etc. – the way in which their interactions are designed, what they may know and not know of each other, how they access the collective procedure, what path their actions follows and how it merges with the actions of others, affects the content of the outcome. Of course this is well known by policy makers, constitution writers and all those who participate into design of a democratic system, or any other system of rules that has to take into account the point of view of the many. But the claim may appear less

---

<sup>1</sup> A version of this article has been presented at a workshop on *Collective Wisdom* held at Collège de France in May 2008. I would like to thank Ariel Colonomos, Jon Elster, Scott Page, Dan Sperber and Ariel Suhamy for their useful comments. A previous and shorter version of this paper has been published in English and French on the French magazine *La vie des idées* at <http://www.laviedesidees.fr>.

evident – or at least in need of a more articulate justification - when it deals with the design of knowledge systems based on the automatic collection of unregulated opinions such as it is the case for the epistemic practices on the Web. Furthermore, the Web has entered our social life as a disruptive technology whose immediate effect was to blow up all the existing legitimate procedures of knowledge access, thus “empowering” its users with a new intellectual freedom, the liberty to produce, access and distribute content in countless new ways. Still, methods of tapping into the wisdom of the crowds on the Web are many and much more clearly differentiated than it is usually acknowledged. In his book on the *Wisdom of Crowds*, James Surowiecki writes about the different designs for capturing collective wisdom. He says: “In the end there is nothing about a futures market that makes it inherently smarter than, say, Google. These are all attempts to tap into the wisdom of the crowd, and that’s the reason they work”. Yet, sometimes the devil is in the details and the way in which the wisdom of crowds is captured makes a huge difference on its outcome and its impact on our cognitive life. The *design question* that is thus central when dealing with these systems is: *How can people and computers be connected so that—collectively—they act more intelligently than any individuals, groups, or computers?*

I will try to go through the details of some of the *collective wisdom systems* that are nowadays used on the Web. I will provide a brief description of the design that underlies each of them. Then, I will argue that these systems work because of their very special way of articulating (1) individual choices and collectively-filtered preferences on one hand and (2) human actions and computer processes on the other. I will then conclude by some epistemological remarks about the role of ranking in our epistemic practices, arguing that the success of the Web as an epistemic practice is due to its capacity to provide not so much a potentially infinite system of information storage, but a giant network of ranking and rating systems in which information is valued as long as it has been already filtered by other people. My modest epistemological prediction is that the *Information Age* is being replaced by a *Reputation Age* in which the reputation of an item – that is how others *value* and *rate* the item - is the only way we have to extract information about it. This *passion of ranking* is a central feature of collective wisdom. James Surowiecki imposes a very illuminating list of conditions on the characterisation of a *wise crowd*. Not any crowd is a wise crowd. In order to avoid well-known phenomena such as group polarization, information cascades, conformism, a group must display certain features that make it a potentially intelligent entity. Surowiecki proposes four main characteristics: 1. *diversity of opinion* (each person should have some private information) 2. *independence* (people’s opinions are not determined by others) 3. *decentralization* (people are able to draw on local knowledge) 4. *aggregation* (presence of mechanisms that turn individual judgements in collective decisions). I’m tempted to add a fifth one that is for me crucial especially in order to “speed up” the collective filtering of information: 5. *presence of a rating device* (each person should be able to produce a rating hierarchy, rely on past ranking systems and make – at least in some circumstances – his or her rating available to other persons).

Here I would like to make a more general point about the role of past evaluations and preferences in filtering information. I'll start with a parallel with some famous remarks Edmund Burke wrote about the importance of traditions. Burke was suspicious of revolutions because they risked to wipe out centuries of *tradition*, that is, of patiently collected and selected values, judgements and preferences refined throughout the ages. And this process of refinement is for Burke the essence of *civilisation*, of this thick cultural lore of judgements, values and opinions that penetrates into our minds through education and socialization and constitutes the necessary background of any form of wise thought. If we do not take into account the lore of traditions, we are condemned for Burke to reinvent the wheel at each generation. Our capacity of thinking the world and the institutions around us is much more limited without the contribution of the preferences already aggregated in the past by others. As he says:

« *We are afraid to put men to live and trade each on his own private stock of reason; because we suspect that this stock in each man is small, and that the individuals would do better to avail themselves of the general bank and capital of nations, and of ages* »<sup>2</sup>

Burke was politically wrong but, in some sense, epistemically right: there is something true in his reactionary remarks, even if their application to the analysis of the French Revolution is wrong for many reasons. One reason why his claims on revolutions are unacceptable today is that obviously not all traditions are worth preserving: the institutional biases and the social pressures that make a political tradition survive may be so wrong and oriented to defend the privileges of just one social class, that it is sometimes wiser to entrust a new generation to rethink the whole institutional design of a society from scratch. But, from an epistemic point of view, he captures the intuition that it is almost impossible to think from scratch, to know from scratch, without taking into account the lore of others' preferences and values as it is filtered by a culture. This is an important epistemological point that evokes a similar, epistemological idea expressed by W.V.O. Quine in a famous article on his mentor, Rudolph Carnap: "*The lore of our fathers is a fabric of sentences. In our hands it develops and changes, through more or less arbitrary and deliberate revisions and additions of our own [...] It is a pale gray lore, black with fact and white with convention. But I have found no substantial reasons for concluding that there are any quite black threads in it, or any white ones.*"<sup>3</sup> That is, the lore of a tradition, even of a scientific tradition, doesn't transmit just a bunch of facts from a generation to another, but a sophisticated ensemble of judgements and conventions that shape the way facts will be extracted and classified in a given culture at a given time.

Preferences, conventions and values that others have expressed thus play a critical role in the making of collective wisdom: they shape the reputational landscape that we use to organize our

---

<sup>2</sup> Cf. E. Burke (1790) *Reflections on the Revolution in France*, in E. Burke, *Works*, London, 1867.

<sup>3</sup> Cf. W. V. O. Quine (1954) « Carnap and Logical Truth » reprinted in W. V. O. Quine (1961) *The Ways of Paradox and Other Essays*, Harvard UP, Cambridge, MA.

own heuristics to extract information and provide a - sometimes reliable and sometimes too biased - shortcut to what is worth keeping, remembering and preserving as knowledge. The epistemological enquiry to collective wisdom I am advocating here implies that reputation and rating systems are an essential ingredient of collective processes of knowledge: their cognitive role in extracting information doesn't depend on the intrusion of social factors than are external to the epistemological process, as many have argued. Reputation is a rational criterion of information extraction, a fundamental shortcut for cumulating knowledge in processes of collective wisdom and an ineludible filter to access facts. In my view, in an information-dense environment, where sources are in constant competition to get attention and the option of the direct verification of the information is simply not available at reasonable costs, evaluation and rankings are epistemic tools and cognitive practices that provide an inevitable shortcut to information. This is especially striking in contemporary informationally-overloaded societies, but I think it is a permanent feature of any extraction of information from a corpus of knowledge. There is no ideal knowledge that we can adjudicate without the access to previous evaluations and adjudications of others. No Robinson Crusoe's minds that investigate and manipulate the world in a perfect solitude. My modest epistemological prediction is that the higher is the uncertainty on the content of information, the stronger is the weight of the opinions of others in order to establish the quality of this content. Even if not all collective-wisdom systems embody the access to others' judgement and rankings, I believe that those which do will rapidly outperform, in terms of reliability, the random aggregation of multiple judgements and preferences on which many collective-wisdom systems are now based, as it is shown by the growing impact of the Web 2.0 (the "Social Web") on our epistemic practices.

Of course, this opens the epistemological question of the *epistemic value* of these rankings, that it, to what extent their production and use by a group changes the ratio between truths and falsities produced by that group and, individually, how an awareness of rankings should affect a person's beliefs. After all, rankings introduce a bias in judgement and the epistemic superiority of a biased judgement is in need of justification. Moreover, these rankings are the result of collective human registered activities with artificial devices. The control of the heuristics and techniques that underlie this dynamics of information may be out of sight or incomprehensible for the users who find themselves in the very vulnerable position of relying on external sources of information through a *dynamic, machine-based* channel of communication whose heuristics and biases are not under their control. For example, that companies used to pay to be included in search engines or gain a "preferred placement" was unknown to 60% of users<sup>4</sup> until the American Federal

---

<sup>4</sup> Princeton Survey Research Associates, "A Matter of Trust: What Users Want from Websites", Princeton, January 2002, at: <http://www.consumerWebwatch.com/news/report1.pdf>. The case is reported in R. Rogers (2004) *Information Politics on the Web*, MIT Press.

Trade Corporation wrote in 2002 a public recommendation asking to search engines companies to disclose paid link policies and clearly mark advertisements to avoid users' confusion.

The epistemic status of these collectively produced rankings thus opens a series of epistemological questions:

1. Why do people trust these rankings and should they?
2. Why should we assume that the collective filtering of preferences produces wiser results on the Web?
3. What are the heuristics and biases of the aggregating systems on the Web that people should be aware of?

These questions include a descriptive as well as a normative perspective on the social epistemology of *collective wisdom systems*. A socio-epistemological approach to these questions - as the one I endorse - should try to elucidate both perspectives. Although here I will explore more the descriptive side of the question, by showing the design of collective wisdom systems with their respective biases, the general epistemological reflections introduced above suggest also a possible line of answer to the normative issues. Ways of taking into account the judgements of others are many, some more reliable than others. They can be maintained for strong historical and cultural reasons - as for example it is the case of the rating systems in very idiosyncratic cultural domains such as wine - <sup>5</sup> and undergo periodical "adjustments" to fit better people's actual preferences and values, or they can be maintained just for rational criteria because they tend to an *epistemic optimality*, that is, they produce a better ratio between truths and falsities than alternative ways of extracting information, as in the case of the PageRank algorithm underlying *Google* searches. I don't think that there are neither "pure" cultural ranking systems deprived of any epistemic value that stabilize independently of any cognitive function (apart from extreme cases such as religious ones which require an incredible amount of public reinforcement to be maintained: the angelic hierarchies don't spread around without a massive amount of predicament about them), nor pure "rational" systems that are not biased by the culture and values of their time (even PageRank has to tell apart what are the *hubs* and the *authorities* of the Web, that is, those special sites "whose" epistemic weight is heavier). And there are not pure gullible subjects who are penetrated by the judgements of others without any personal filtering, nor pure rational beings whose heuristics are not biased by their cultural background and their values. We all use these socially constructed information to navigate knowledge, acquire competence in new domains, and balance what is worth keeping and what is worth challenging by the epistemic use we make of these devices. This doesn't make us more gullible. Our epistemic responsibility in dealing with these reputational devices is to be aware of the biases that the design of each of these devices incorporates, either for technical reasons or for sociological,

---

<sup>5</sup> Cf. G. Origgi (2007) « Wine epistemology : The role of réputation and rating systems in the world of wine » in B. Smith (ed.) *Questions of Taste. The Philosophy of Wine*, Oxford UP.

historical or even institutional reasons. A detailed presentation of what sort of aggregation of individual choices the Internet makes available should be thus accompanied by an analysis of the possible biases that each of these systems carries in its design. The awareness of these biases is part of our epistemic competence, as the awareness of the limits and biases of our perceptual mechanisms should be part of our individual epistemic competence in extracting information from the environment around us. But let us explore the design of some of these systems of collective intelligence, the role of the ranking devices that they incorporate and the specific biases that we should be aware of.

### *Collective intelligence out of individual choices*

People - and other intelligent agents - often think better in groups and sometimes think in ways which would be simply impossible for isolated individuals. The Internet is surely an example of this. That is why the rise of the Internet created from the onset huge expectations about a possible “overcoming” of thought processes at the individual level, towards an emergence of a new – more powerful – form of technologically-mediated intelligence. A plethora of images and metaphors of the Internet as a super-intelligent agent thus invaded the literature on media studies – such as the Internet as an extended mind, a distributed digital consciousness, a higher-order intelligent being, etc...

Yet, the collective processes that make Internet such a powerful cognitive media are precisely an example of “collective intelligence”, that is, a mean of aggregation of individual choices and preferences. What Internet made possible though – and this was indeed spectacular - was a brand new form of aggregation that simply didn’t exist before its invention and diffusion around the world. In this sense, it provided a new tool for aggregating individual behaviours that may serve as a basis for rethinking other forms of institutions whose survival depends on combining in the appropriate way the views of the many.

#### *1.1. The Internet and the Web*

The salient aspect of this new form of aggregation is a special way of articulating individual choices and collectively-filtered preferences through the technology of the Internet and, especially, of the World Wide Web. In this sense, it is useful to distinguish from the onset between the Internet as a networking phenomenon and the Web as a specific technology made possible by the existence of this new network. The Internet is a network whose beginnings go back to the Sixties, when American scientists at AT&T, Rand and MIT and the Defense Communication Agency started to think of an alternative model of transmitting information through a network. In the classical telephone system, when you call New York from your apartment in Paris, a circuit is open between you and the New York destination – roughly a copper line which physically connects the two destinations. The idea was thus to develop an

alternative – “packet-switching” technology, by digitalizing conversations – that is – translating waves into bits, then chopping the result into packets which could flow independently through a network while giving the impression of a real-time connection on the other end. In the early Seventies the first decentralised network, Arpanet, was put in use that was able to transfer a message by spreading its chunks through the network and then reconstructing it at the end. By the mid Seventies, the first important application on the network, the mail, was created. What made this net such a powerful tool was its decentralised way of growing: Internet is a network of networks, which uses pre-existing wires (like telephone networks) to make computers communicate through a number of protocols (things like: IP/TCP) that are not proprietary: each new user can connect to the network by using these protocols. Each invention of an application, a mail system, a system of transfer of video, a digital phone system, can use the same protocols. Internet protocols are “commons”<sup>6</sup>, and that was a boost to the growth of the network and the creativity of the applications using it. This is a crucial for the wisdom of the net. Without the political choice to keep these protocols free, the net would not have grown in a decentralised manner and the collaborative knowledge practices that it has realized would not have been possible. The World Wide Web, which is a much more recent invention, maintained the same philosophy of open protocols compatible with the Internet (like HTTP –hypertext transfer protocol or HTML- hypertext markup language). The Web is a service which operates through the Internet, a set of protocols and conventions that allows “pages” (i.e. a particular format of information that makes easy to write and read content) to be easily linked to each other, by the technique of hyperlink. It’s a visualization protocol that makes the display of information very simple. The growth of the Web is not the same thing as the growth of Internet. What made the Web grow so fast is that the creating a hyperlink doesn’t require any technical competence. The Web is an illustration of how an Internet application may flourish thanks to the openness of the protocols. And it is true that impact of IT on collective intelligence are due mostly to the Web.

### *1.2. The Web, collective memory and meta-memory*

What makes the aggregation of individual preferences so special through the Web? For the history of culture, the Web is a major revolution on the storage, dissemination and retrieving of information. The major cultural revolutions in the history of culture have had an impact on the distribution of memory, and Web is one such revolution. Let’s see in what sense. The Web has often been compared to the invention of writing or printing. Both comparisons are valid. Writing, introduced at the end of the 4th millennium BCE in Mesopotamia, is an external memory device that makes possible the reorganization of intellectual life and the structuring of thoughts, neither of which are possible in oral cultures. With the introduction of writing, one part of our cognition “leaves” the brain to be distributed among external supports. The visual

---

<sup>6</sup> Cf. on this point, L. Lessig (2001) *The Fututre of Ideas*, Vintage, New York.

representation of a society's knowledge makes it possible to both reorganize the knowledge in a more useful, more 'logical', way by using, for example, lists, tables, or genealogical trees, and to solidify it from one generation to the next. What's more, the birth of "managerial" casts who oversee cultural memory, such as scribes, astrologists, and librarians, makes possible the organization of meta-memory, that is, the set of processes for accessing and recovering cultural memory.

Printing, introduced to our culture at the end of the 15th century, redistributes cultural memory, changing the configuration of the informational pyramid in the diffusion of knowledge. In what sense is the Web revolution comparable to the invention of writing and printing? In line with these two earlier revolutions, the Web increases the efficiency of recording, recovering, reproducing and distributing cultural memory. Like writing, the Web is an external memory device, although different in that it's "active" in contrast to the passive nature of writing. Like printing, the Web is a device for redistributing the cultural memory in a population, although importantly different since it crucially modifies the costs and time of distribution. But unlike writing and printing, the Web presents a radical change in the conditions for accessing and recovering cultural memory with the introduction of new devices for managing meta-memory, i.e., the *processes for accessing and recovering memory*. Culture, to a large extent, consists in the conception, organization and institutionalization of an efficient meta-memory, i.e. a system of rules, practices and representations that allow us to usefully orient ourselves in the collective memory. A good part of our scholastic education consists in internalizing systems of *meta-memory*, classifications of style, rankings, etc.. chosen by our particular culture. For example, it's important to know the basics of rhetoric in order to rapidly "classify" a line of verse as belonging to a certain style, and hence to a certain period, so as to be able to thus efficiently locate it from within the corpus of Italian literature. Meta-memory thus doesn't serve only a cognitive function – to retrieve information from a corpus – but a social and epistemic function to provide an organization for this information in terms of various systems of classifications that embody the value of the "cultural lore" of that corpus. The way we retrieve information is an epistemic activity which allows us to access through the retrieving filters, how the cultural authorities on a piece of information have classified and ranked it within that corpus. With the advent of technologies that automate the functions of accessing and recovering memory, such as search engines and knowledge management systems, meta-memory also becomes part of external memory: a cognitive function, central to the cultural organization of human societies, has become automated—another "piece" of cognition thus leaves our brain in order to be materialized through external supports. Returning to the example above, if I have in mind a line of poetic verse, say "Guido, i'vorrei..." but can recall neither the author nor the period, and am unable to classify the style, these days I can simply write the line of verse in the text window of a search engine and look at the results. The highly improbable combination of words in a line of

verse makes possible a sufficiently relevant selection of information that yields among the first results the poem from which the line is taken.

How is this meta-memory designed through the Web technology? What is unique on the Web is that the actions of the users leave a track on the system that is immediately reusable by it, like the trails that snails leave on the ground, which reveal to other snails the path they are following. The combination of the tracks of the different patterns of use may be easily displayed in a rank that informs and influence future preferences and actions of the users. The corpus of knowledge available on the Web – built and maintained by the individual behaviours of the users – is automatically filtered by systems that aggregate these behaviours in a ranking and make it available as filtered information to new, individual users. I will analyse two different classes of meta-memory devices. These systems, although they both provide a selection of information that informs and influences users' behaviour, are designed in a different way, a difference is worth taking notice of.

## 2. Collaborative filtering: wisdom out of algorithms

### 2.1. Knowledge Management Systems

Collaborative filtering is a way of making predictions about the preferences of a users based on the pattern of behaviour of many other users. It is mainly used for commercial purposes in web applications for e-business, although it has been extended to other domains. A well-known example of a system of collaborative filtering which I assume we are all familiar with, is Amazon.com : Amazon.com is a Web application, a knowledge management system which keeps track of users' interactions with the systems and is designed to display correlations between patterns of activities in a way that informs users about other users' preferences. The best known feature of this system is the one which associates different items to buy: *Customers who buy X buy also Y*. The originality of these systems is that the matching between X and Y is in a sense *bottom-up* (although the design of the appropriate thresholds of activities above which this correlation emerges are fixed by the information architecture of the system). The association between James Surowiecki's book and Ian Ayer's book *Super Crunchers* that you can find on the Amazon's page for *The Wisdom of Crowds* has been produced automatically by an algorithm that aggregates the preferences of the users and makes the correlation emerge. This is a unique feature of these interactive systems, in which new categories are created by automatically transforming human actions into visible rankings. The collective wisdom of the system is due to a division of cognitive labour between the algorithms which compose and visualize information, and the users who interact with the system. The classifications and rankings that are thus created aren't based on previous cultural knowledge of habits and customs of users, but on the emergence of significant patterns of aggregated preferences through the individual interactions with the system. Of course, biases are possible within the system: the weights associated to each item to make it

emerge are fixed in such a way that some items have more chances to be recommended than others. But given that the system is alimented by the repeated actions of the users, a too biased recommendation that couples items that users won't buy together will not be replicated enough times to stabilize within the system.

## 2.2. PageRank

Another class of systems that realize meta-memory functions through artificial devices are search engines. As we all know by experience, search engines have been a major transformation of our epistemic practices and a profound cognitive revolution. The most remarkable innovation of these tools is due to the discovery of the structure of the Web at the beginning of this century<sup>7</sup>. The structure of the Web is that of a social network, and contains a lot of information about its users' preferences and habits. The search engines of second generation, like *Google*, are able to exploit this structure in order to gain information about how knowledge is distributed throughout the world. Basically, the PageRank algorithm interprets a link from a page A to page B as a vote that page A expresses towards page B. But we're not in democracy on the Web and votes do not have all the same weight. Votes that come from certain sites – called “hubs”- have much more weight than others, and reflect in a sense hierarchies of reputation that exist outside the Web. Roughly, a link from my homepage to the page of the Harvard University weighs much less than a link to my page from that of Harvard University. The Web is an “aristocratic” network – an expression that is used by the social network theorists – that is, a network in which “rich get richer” and the more links you receive the higher is the probability that you will receive even more. This disparity of weights creates a “reputational landscape” that informs the result of a query. The PageRank algorithm is nourished by the local knowledge and preferences of each individual user and it influences them by displaying a ranking of results that are interpreted as a hierarchy of relevance. Note that this system is not a knowledge management system: The PageRank algorithm doesn't know anything about the particular pattern of activities of each individual: it doesn't know how many times you and I go to the JSTOR website and doesn't combine our navigation paths together. A click from a page to another is an opaque information for PageRank, whereas a link between two pages contains a lot of information about users' knowledge that the system is able to extract. Still, the two systems are comparable from the point of view of the design of collective intelligence: neither requires any cooperation between agents in order to create a shared system of ranking. The collaborative aspect of the collective filtering

---

<sup>7</sup> Kleinberg, J. (2001) “The Structure of the Web”, *Science*, vol. 294, pp. 1849-50.

is more in the hands of machines than of human agents<sup>8</sup>. The system exploits the information that human agents either unintentionally leave on the website by interacting with it (KM systems) or actively produce by putting a link from one page to another (search-engines): the result is collective, but the motivation is individual.

Biases of search engines have been a major subject of discussions, controversy and collective fears these years. As I have mentioned above, the refinement of the second-generation search engines such as Google has allowed at least to explicitly mark paid inclusions and preferred placements, but this needed a political intervention. Also, the *Mathiew effect* of aristocratic networks is notorious, and the risk of these tools is to give prominence to already powerful sites at the expense of others<sup>9</sup>. The awareness of these biases should imply a refinement on the search practices also: for example, the more improbable is the string of keywords, the more relevant is the filtered result. Novices and learners should be instructed with even simple principles that make them less vulnerable to these biases.

### 3. Reputation systems: wisdom out of status anxiety

The collaborative filtering of information may require sometimes a more active participation to a community than what is needed in the examples above. In his work on *Information Politics on the Web* the sociologist Richard Rogers classifies web dynamics as *voluntaristic* or *non-voluntaristic* according to the respective role of human and machines in providing information feed-back for the users. *Reputation systems* are an example of a more voluntaristic web application than the ones seen above. A reputation system is a special kind of collaborative filtering algorithm that determines ratings for a collection of agents based on the opinions that these agents hold about each other. A reputation system collects, distributes, and aggregates feedback about participants' past behaviour.

The best known and probably simplest reputation system of large impact on the Web is the system of auction sales at [www.eBay.com](http://www.eBay.com). eBay allows commercial interactions among more than 125 millions of people around the world. People are buyers and sellers. Buyers place a bid on an item. If their bid is successful, they make the commercial transaction, then both (buyers

---

<sup>8</sup> Knowledge management systems like Amazon.com have some collaborative filtering features that need cooperation, like writing a review of a book or ranking a book with the five stars ranking system, but these aren't essential to the functioning of the collaborative filtering process.

<sup>9</sup> The expression *Mathiew effect* was coined by the sociology Robert K. Merton in his work in sociology of science, where he claimed that credit is always given in an unfair way to those who have it already. The expression refers to the famous passage in the Gospels : « For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath », Mathiew, 25 :29.

and sellers) leave a feedback about the quality of that transaction. The different feedbacks are then aggregated by the system in a very simple feedback profile, where positive feedbacks and negative feedbacks plus some comments are displayed to the users. The reputation of the agent is thus an useful information in order to decide to pursue the transaction. Reputation has in this case a real, measurable, commercial value: in a market with a fragmented offer and very low information available on each offer, reputation becomes a crucial information in order to trust the seller. Sellers on eBay know very well the value of their good reputation in such a special business environment (no physical encounters, no chance to see and touch the item, vagueness about the normative framework of the transaction – if for example it is realized through two different countries, etc.), so there is a number of transactions at a very low cost whose objective is just to gain one more positive evaluation. The system creates a collective result forcing cooperation, that is, asking users to leave an evaluation at the end of the transaction and sanctioning them if they don't comply. Without users' active participation, the system will be useless. Still, it is a special form of collaborative behaviour that doesn't require any commitment to cooperation as a value. Non-cooperative users are sanctioned to different degrees: they can be negatively evaluated not only if the transaction isn't good, but also if they do not participate into the evaluation process. Breaking the rules of e-bay may lead to the exclusion from the community. The design of wisdom thus comprises an active participation from the users for fear to be ostracized by the community (which would be seen as a loss of business opportunities). Biases are clearly possible here also. People invest in cheap transactions whose only aim is to gain reputational points. This is a bias one should be aware of and easily check: if a seller offers too many cheap items, he too concerned with his public image to be considered reliable. Some reputational features are used also by non-commercial systems such as [www.flickr.com](http://www.flickr.com). Flickr is a collaborative platform to share photos. For each picture, you can visualise how many users have added it among their favourite pictures and who they are. Reputation systems differ from other systems of measurement of reputation that use citation analysis, like for example the *Science Citation Index*. These systems are in a sense reputation-based, given that they use scientometric techniques to measure the impact of a publication in terms of the number of citations in other publications. But they don't require any active participation of the agents in order to obtain the measure of reputation. Reputation is in this latter case established by past rankings and can't be influenced by users' preferences and behaviours<sup>10</sup>. But we can easily imagine an evolution of these systems that will take into account also new measures of reputation, for example judgements of value expressed by the readers of the scientific articles. Many ongoing Web projects aimed at reforming the status of reputation systems of academic

---

<sup>10</sup> On *Citation Index*, see G. Origi (2007) « Un certain regard. Pour une épistémologie de la réputation »,

impact based on citation analysis argue that the social dimension of evaluation should be included in the credit attribution to each scientific paper.<sup>11</sup>

#### 4. Collaborative, open systems: wisdom out of cooperation

The collaborative filtering on the Web may be even more volountaristic and human-based than the previous examples, while still necessitating a Web support to realize an intelligent outcome. Two are the most discussed cases of collaborative systems that owe their success to active human cooperation in filtering and revising the information made available: the Open Source communities of software development, like for example the community around the development of the operating system *Linux*, and the collective open content projects such as *Wikipedia*. In both cases, the filtering process is completely human-made: code or content is made available to a community which can filter it by correcting, editing or erasing it according to personal or shared standards of quality. I would say that these are communities of *amateurs* instead of experts, that is, people who love what they do and decide to share their knowledge for the sake of the community. Collective wisdom is thus created by individual human efforts that are aggregated in a common enterprise in which some norms of cooperation are shared.

I won't discuss biases on Wikipedia: it is such a large topic that it could be the subject of another paper. Let me just mention that Larry Sanger, one of its founders, is promoting an alternative project, [www.citizendum.org](http://www.citizendum.org) which endorses a policy of accreditation of its authors. Self-promotion, ideology, targeted attacks on reputation may of course act as biases in the selection of entries. But the fear of Wikipedia as a dangerous place of tendentious information has been disconfirmed by facts: thanks to its large size, Wikipedia is hugely differentiated in its topics and views, and it has been shown that its reliability is no less than that of the Encyclopedia Britannica,<sup>12</sup> although the validity of *Nature's* study on Wikipedia has been challenged by more recent studies<sup>13</sup>. It is true that the democracy and size are not the only ingredients that assure reliability: Wikipedia is a community that works thanks to the "fanatism" of a group of members who are committed to its values and aims and actively participate into its maintenance and

---

<sup>11</sup> See for example the new European Commission project *Liquidpublication* : <http://wiki.liquidpub.org/> whose aim is to promote an alternative Web-based system of scientific production that takes into account crédits established by the scientific community as well as crédits cumulated through the opinions expressed by the actual readers of the articles.

<sup>12</sup> Cf. "Internet Encyclopedias go head to head" *Nature*, 438, 15 December 2005.

<sup>13</sup> Cf. The reply to L. Giles *Nature* enquiry written by the Encyclopedia Britannica at : [http://corporate.britannica.com/britannica\\_nature\\_response.pdf](http://corporate.britannica.com/britannica_nature_response.pdf), and Nicolas Carr's critical analysis of the results at : [http://www.rougtype.com/archives/2006/03/britannicas\\_ind.php](http://www.rougtype.com/archives/2006/03/britannicas_ind.php).

quality control. As Florence Devouard, the ex-president of the French Wikimedia Foundation, recently pointed out<sup>14</sup>, Wikipedia works because it is organized around a multiplicity of different communities (usually sorted by language, nation, or focus of interest) each of which is strongly animated by a group of so-called “glue-people” or “bridge-people” that allow to share the best practices among the members of a community and to transfer the most efficient practices from one community to another. Also, Wikipedia works because it deals with sharing cultural information, which diminishes the size of potential contributors (people who are able to write articulated entries on Wikipedia are a much less extended community than people who are able to sell their products on the Web), and because cultural information may be “chopped” in small pieces: Each contributor may just contribute for a small chunk of information so that an entry’s content increments slowly by adding piece after piece. Rules and norms of cooperation thus play a major role in the growing and maintenance of this collective wisdom system.

### *Sharing folksonomies: wisdom out of tagging information*

As we have seen up to now, the most interesting and promising collective wisdom systems nowadays are *social computing systems*, that is, social software that supports the desire of individuals to be pulled into groups to achieve goals. Among these, systems based on social tagging, that is, those which categorize information on the basis on the spontaneous categorizations that users produce, are among the best aggregators of collective wisdom. « Tagging » is the act of adding keywords, also known as tags, to any type of digital resource. Tags serve to describe the item and enable a keyword-based classification. But they can also be used to search for content. The types of content that can be tagged varies from: blogs (Technorati), books (Amazon), pictures (Flickr), podcasts (Odeo), videos (YouTube), Web pages (Delicious) to even tagging of tags. Tags are not only metadata, but also content. Tagging also allows social groups to form around similarities of interests and points of view, hence the term *social tagging*. Social tagging is one of the web 2.0 success stories, tapping into the 'wisdom of crowds' - it lets users connect with others, enabling social discovery and connections<sup>15</sup>. An example of social tagging system is [www.CiteUlike.org](http://www.CiteUlike.org), a free online service to share the academic papers you are reading with other people in your scientific community, find out who is reading the same papers as you, organize a library of some of the papers you are reading by tagging them in a particular way. Preferences are here voluntary shared. Some tagging systems collect information from users by actively asking them to rate a number of items, or to express a preference between two items, or to create a list of items that they like. It is basically a collaborative filtering technique with a more active component: people

---

<sup>14</sup> Cf. Florence Devouard’s speech at the *World Knowledge Dialogue* (Crans, Switzerland, Sept. 2008) : <http://www.wkdnews.org/webcast?clip=9>.

<sup>15</sup> Cf. G. Veltri, « Social Computing and Social Science », TGE Adonis CNRS report, December 2008.

are asked to express their preferences, instead of just inferring their preferences from their behaviour, which makes a huge difference: it is well known in psychology that we are not so good in introspection and sometimes we consciously express preferences that are incoherent with our behaviour: If asked, I may express a preference for classical music, while if I keep a record of how many times I do listen to classical music compared other genres of music in a week, I realize that my preferences are quite different. Folk taxonomies produced by social tagging are exactly “traditions” in the Burkean sense I’ve tried to clarify at the beginning of the article: they provide filtered, evaluated information that we may use to empower our own way of filtering and classifying the world. The difference is that these traditions are instantaneously produced and not institutionally reinforced: they can be created and recreated anytime and if they last it is because they serve a certain cognitive function and not an ideological one. And the more information is shared and tagged, the more it acquires value for further sharing and tagging, in a sort of “cornucopia of the commons” in which use brings overflowing abundance.<sup>16</sup>

### *Conclusions*

This long list of examples of Web tools for producing collective wisdom illustrates how fine-grained can be the choice of the design for aggregating individual choices and preferences: “Architecture is politics” is the slogan nowadays of the promoters of the social Web. The differences in design that I have underlined end up in deep differences in the kind of collective communities that are generated by the IT. Sometimes the community is absent, as in the case of the Google users, who cannot be defined as a “community” in any interesting normative sense, sometimes the community is normatively demanding, as in the case of eBay, in which participation in the filtering process is needed for the survival of the community. If the new collective production of knowledge that the Web – and in particular the Web 2.0 and 3.0 – makes possible should serve as a laboratory for designing “better” collective procedures for the production of knowledge or of wise decisions, these differences should be taken into account. But let me come back in the end with a more epistemological claim about what kind of knowledge is produced by these new tools. As I said at the beginning, these tools work insofar as they provide access to rankings of information, labelling procedures and evaluations. Even Wikipedia, which doesn’t display any explicit rating device, works on the following principle: if an entry has survived on the site – that is, it has not been erased by other wikipedians – it is worth reading it. This can be a too weak evaluative tool, and, as I said, discussion goes on these days on the opportunity to introduce more structured filtering devices on Wikipedia<sup>17</sup>, but it is

---

<sup>16</sup> On the idea of « cornucopia of the commons », that echoes by contrast the famous G. Hardin’s paper on « The Tragedy of the Commons », *Science*, 1968, vol. 162, n. 3859, pp. 1243-1246, see Dan Bricklin’s blog: <http://www.bricklin.com/cornucopia.htm>.

my opinion that the survival or even egalitarian projects like Wikipedia depends on their capacity to incorporate a ranking: the label *Wikipedia* in itself works already as a reputational cue that orients the choices of the users. Without the reputation of the label, the success of the project would be much more limited.

As I said at the beginning, the Web is not only a powerful reservoir of all sort of labelled and unlabelled information, but it is also a powerful *reputational* tool that introduces ranks, rating systems, tags, weights and biases in the landscape of knowledge. Even in this information-dense world, knowledge without evaluation would be a sad desert landscape in which people would be stunned in front of an enormous and mute mass of information, as *Bouvard et Pécuchet*, the two heroes of Flaubert's famous novel, who decided to retire and to go through every known discipline without, in the end, being able to learn anything. An efficient knowledge system will inevitably grow by generating a variety of evaluative tools: that is how culture grows, how traditions are created. A cultural tradition is to begin with a labelling system of insiders and outsiders, of who stays on and who is lost in the magma of the past. The good news is that in the Web era this inevitable evaluation is made through new, collective tools that challenge the received views and develop and improve an innovative and democratic way of selection of knowledge. But there's no escape from the creation of a canonical—even if tentative and rapidly evolving—corpus of knowledge.

## References

- A. Clark (2003) *Natural Born Cyborgs*, Oxford University Press.
- L. Lessig (2001) *The Future of Ideas*, Vintage, New York
- G. Origgi (2007) “Wine epistemology: The role of reputation and rating systems in the world of wine”, in B. Smith (ed.) *Questions of Taste*, Oxford University Press.
- G. Origgi (2007) « Un certain regard. Pour une épistémologie de la réputation », presented at the workshop *La réputation*, Fondazione Olivetti, Rome, April 2007.
- G. Origgi (2008) *Qu'est-ce que la confiance*, VRIN, Paris.
- R. Rogers (2004) *Information Politics of the Web*, MIT Press
- L. Sanger (2007) “Who says we know: On the new Politics of knowledge” at [www.edge.org](http://www.edge.org)
- Taraborelli, D. (2008) “How the Web is changing the way we trust”, in: K. Waelbers, A. Briggel, P. Brey (Eds.), *Current Issues in Computing and Philosophy*, IOS Press, Amsterdam, 2008.
- C. R. Sustain (2006) *Infotopia*, Oxford University Press.
- P. Thagard (2001). Internet epistemology: Contributions of new information technologies to scientific research. In K. Crowley, C. D. Schunn, and T. Okada, (Eds.) *Designing for science: Implications from professional, instructional, and everyday science*. Mahwah, NJ: Erlbaum, 465-485.

---

<sup>17</sup> See. L. Sanger «”Who says we know. On the new politics of knowledge” on line at [www.edge.org](http://www.edge.org) and my reply to him, G. Origgi “Why reputation matters”.

