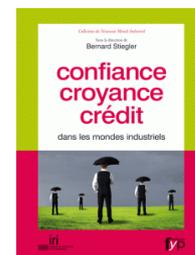


## Intelligence collective et confiance démocratique Nicolas Auray

in *Confiance, croyance, crédit dans les mondes industriels*, B.Stiegler (éd.), 2012, FYP Editions



La culture et la connaissance baignent dans une tension explosive. D'une part, selon des dynamiques *prédatoriales*, pionniers audacieux, entrepreneurs conquérants et autres créateurs de tendance, exploitant les externalités de réseau du capitalisme dit « cognitif », deviennent des oligopoles – majors, grands éditeurs, consoliers. Ils s'arrogent des positions dominantes, colonisent la distribution, parasitent la prescription<sup>1</sup> et forment des rentes artificielles. En opposition à ces conglomerats qui capturent l'attention d'un grand public volage, s'est émancipée – d'abord autour de *l'open source*<sup>2</sup> puis du web participatif – une dynamique *exploratoire* venue de collectifs de consommateurs et d'usagers, voulant répartir plus également la visibilité entre un plus grand nombre d'œuvres produites, mutualisant leurs avis dans des *communautés* en ligne d'amateurs, et donnant une très large place aux contenus rares et aux nouveaux talents. Un tel web de la recommandation n'est pas exempt lui-même de mouvements de et de logiques prédatrices, mais il oppose au mimétisme et à la viralité un régime de la « rétention »<sup>3</sup> et de l'attention patiente, fondé sur *l'exploration curieuse*, où cohabitent l'envie d'être surpris et la focalisation sur un intérêt thématique. Ces collectifs explorateurs s'inscrivent dans la tradition des amateurs au sens où ils assimilent l'expérience esthétique à l'apprentissage de langages critiques, apprenant à sentir les œuvres d'art par les langages dans lesquelles elles se disent.

Ancêtres de ce régime exploratoire, et rebelles au contrôle excessif exercé sur la détection de talents par les conglomerats médiatiques, les premiers blogueurs se définissaient comme des *trailblazers*, des éclaireurs<sup>4</sup>. Orpailleurs de la connaissance, ils s'inscrivaient, lorsqu'ils lancèrent la forme « blog » comme « trajectoire commentée de pages méconnues », dans l'héritage de l'article « As we may think » de l'ingénieur Vannevar Bush ou du livre *Langage, Thought, and Reality* du linguiste Benjamin Lee Whorf<sup>5</sup>. De même, les communautés du logiciel libre se sont structurées autour d'une commune valorisation du code informatique ouvert comme recueil encyclopédique de trouvailles d'écriture et de méthodes de programmation, et les communautés de pratique constituent

<sup>1</sup> Cf. Richard E. Caves, *Creative Industries : Contracts between Art and Commerce*, Harvard University Press, 2000 ; Sherwin Rosen, « The Economics of Superstars », *American Economic Review*, 1981, vol. 71, n° 5, p. 845-858 ; Philippe Chantepie et Alain Le Diberder, *Révolution numérique et industries culturelles*, La Découverte, 2006.

<sup>2</sup> Le terme qualifie, au-delà du monde du « logiciel libre », des contenus redistribuables librement et dont les travaux dérivés sont en accès libre.

<sup>3</sup> Certes, toute technologie intellectuelle, en tant que support de mémoire, est une « rétention tertiaire », comme le dit Bernard Stiegler. L'originalité des communautés virtuelles est d'identifier la rétention à l'apprentissage mutuel – par intégration du point de vue d'autrui et différenciations mutuelles. Les jeux de langage permettent de développer le jugement de goût en « expérience esthétique ». Comme l'a montré Yves Michaud par exemple, à la suite du philosophe Ludwig Wittgenstein, dans ce que nous appelons les arts on commence par un jugement rudimentaire (« cool », « super », « vachement bien », « chouette », « magnifique », etc.) puis on développe en fonction de son insertion dans le jeu de langage.

<sup>4</sup> Cf. Torill Mortensen et Jill Walker, « Blogging thoughts : personal publication as an online research tool », in Morrison Andrew, *Researching Icts in Context*, InterMedia/UniPub, 2002.

<sup>5</sup> Cf. Geert Lovink, *Zero Comments : Blogging and critical Internet culture*, Routledge, 2006.

des creusets d'apprentissage d'un thème commun à partir de collections variées<sup>6</sup>. Les sites de folksonomy c'est-à-dire d'indexation à partir de « tags », permettent des découvertes inattendues, en cherchant l'optimum magique entre l'étonnement recherché et le plaisir continué<sup>7</sup>. Contrairement à la recherche d'information par requête à un moteur de recherche, marquée par l'usage *instrumental* d'un logiciel pour limiter à une réponse précise une investigation déjà formatée, l'internet culturel – web 2.0, plateformes participatives, communautés – s'appuie sur des techniques aléatoires qui soutiennent une exploration *désordonnée* et *curieuse*<sup>8</sup>.

Plusieurs traits caractérisent ces collectifs informationnels, dont l'ouverture est relancée par le numérique: le goût pour l'énigme, la farce et le canular; l'usage d'instruments qui maximisent l'écart entre une variation minimale des conditions initiales et les conséquences imprévues; la valeur donnée à la capacité, appelée « sérendipité », à faire feu de tout bois. Pour résumer d'un mot ce nouveau régime d'attention qu'ils construisent, j'ai proposé le terme de « fascination curieuse », qui renvoie à la menace d'une addiction alors même que l'activité est marquée par le tâtonnement et l'enquête<sup>9</sup>.

L'objet de ce texte est de mieux comprendre comment se stabilisent ces communautés originales, qui organisent l'accès à un corpus toujours extensible de connaissances et de biens. Comment se protègent-elles contre les vandales, les spammeurs et les tricheurs? Comment s'y construisent et s'y articulent trois confiances: la confiance envers les autres contributeurs, souvent inconnus, la confiance dans la connaissance et sa valeur, la confiance en soi ou l'assurance de ne pas être crédule? Comment éviter que ces régulations ne fassent sombrer le collectif dans des tyrannies de la visibilité et de la surveillance participative?

Une originalité forte de la gouvernance de ces collectifs doit tout d'abord être soulignée: la permissivité. Elle part du principe que s'y réalise une « inversion de la tragédie des communs »<sup>10</sup>: alors que les communautés du monde réel, organisées autour de la protection d'une ressource commune, protègent l'écologie de ces écosystèmes fragiles – bassins de pêche, systèmes d'irrigation, etc. – par des *sanctions* envers les inciviques, certes discutées et graduées<sup>11</sup>, les communautés ouvertes sont presque toujours *permissives* aux apports, car leur inquiétude porte non pas sur l'asphyxie par des pollueurs mais sur la menace de sous-alimentation en contenus. Elles pratiquent le *signalement d'erreur* plus que la sanction ou le blâme, et séparent soigneusement l'évaluation des contenus et les jugements sur les personnes. Dès lors, elles fabriquent les bases d'une confiance démocratique, dans laquelle s'articulent des mécanismes de création de la confiance indépendants de l'autorité sociale des personnes, mécanismes fondés sur des délibérations souveraines qui en construisent de manière transparente les procédures. Nous distinguerons d'abord deux dimensions du principe de l'ouverture. L'ouverture des *frontières* de la communauté des contributeurs permet à ces collectifs d'atteindre des tailles gigantesques en étant disponibles à des apports passagers de membres non inscrits. Elle amène à une conception centrée sur le contenu des calculs de confiance. L'ouverture des *statuts* des membres, révisables, stimule les apports individuels en attribuant des rangs proportionnels à la qualité des contributions telle qu'elle est évaluée par leurs pairs. Elle introduit aussi à des techniques sophistiquées pour parer la crainte que des utilisateurs malintentionnés fassent un usage stratégique de leur connaissance des règles de

---

<sup>6</sup> Cf. Etienne Wenger, *Communities of Practice. Learning, Meaning, and Identity*, Cambridge University Press, 1998.

<sup>7</sup> Les algorithmes de soutien à la recommandation sur ces sites sont spécifiquement construits pour échapper à la régression vers une opinion majoritaire (tout en évitant la babélisation). Par exemple, dans le site Delicious, pour aider les gens à attribuer des mots-clés aux liens web qu'ils mettent en partage, un suggestionneur de tags leur propose le cas échéant uniquement l'intersection entre l'ensemble des tags déjà marqués pour cet item et l'ensemble des tags déjà mis par la personne. Le but est d'assurer la lisibilité tout en maintenant élevée la singularité, la différenciation des façons d'indiquer, pour que celles-ci continuent à bien porter des représentations variées.

<sup>8</sup> Cf. Nicolas Auray, « Folksonomy : a New Way to Serendipity », *Communications & Strategies*, 2007, n° 65, p. 67-91.

<sup>9</sup> Cf. Nicolas Auray, « Manipulation à distance et fascination curieuse : les pièges liés au spam », *Réseaux*, 2012, n° 171, p. 104-131; et « Les technologies de l'information et le régime exploratoire », in Danièle Bourcier et Pek Van Aniel (dir.), *La Sérendipité : le hasard heureux*, Hermann, 2011.

<sup>10</sup> Cf. Garrett Hardin, « Extensions of "The Tragedy of the Commons" », *Science*, 1998, vol. 280, n° 5364, p. 682-683.

<sup>11</sup> Cf. Elinor Ostrom, *Governing the Commons : The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.

construction des indicateurs de statut. Puis, après avoir fait cette distinction, nous étudierons les modalités par lesquelles ces collectifs se protègent des implications tyranniques liées à la surveillance mutuelle exercée sur les contenus.

## 1- « **Be bold, newcomers** »<sup>12</sup> : la confiance aux frontières

Suivant le principe que, dans la communauté ouverte, les internautes veulent butiner de proche en proche de manière à découvrir par heureux hasard des contenus qu'ils n'auraient pas recherchés au départ<sup>13</sup>, la plupart des gros collectifs en ligne se caractérisent par l'accueil des contributions opportunistes et ponctuelles. Ils sont ouverts aux « coopérations faibles »<sup>14</sup>. Ainsi, les *user lists* (listes des utilisateurs) ouvrent un cadre d'intelligibilité qui rend pertinents les énoncés les plus inattendus ou surprenants. Elles sont ouvertes en écriture même à des utilisateurs non abonnés. L'ouverture curieuse est valorisée pour ses bénéfices épistémiques : elle permet de construire les conditions favorables à la découverte. Les initiatives pour restreindre cette ouverture sont subordonnées à l'acceptation par une majorité qualifiée. La fragmentation des listes est un processus strictement contrôlé. Sur Usenet, modèle historique des forums de discussion, pour être acceptée, une scission de forum doit « avoir plus de 80 OUI de plus que de NON et plus de 3 fois plus de OUI que de NON ». Seuls de rares messages font l'objet d'un rejet, selon une procédure adoptée par consensus. Usenet utilise l'« indice de Breidbart », qui fait la somme des forums dans lesquels le message a été publié, et des racines carrées des forums dans lesquels il a fait l'objet d'un *forward*, une transmission, pour calculer le seuil à partir duquel un message d'utilisateur est interdit sur une liste. Ce seuil est fixé par un consensus entre les membres : en 2010, sur le Big8 (les 8 grandes hiérarchies modérées), un robot rejetait un message si cet indice dépassait 20 sur l'intervalle des 45 derniers jours ; le seuil était ramené à 10 sur 30 jours dans la hiérarchie française.

Une place prioritaire est donnée aux *anonymes* – 65 % des modifications (*edits*) de Wikipédia anglophone et 75 % de la branche francophone sont en 2010 effectuées par des « IP », qui n'ont pas donné leur nom d'état civil, n'ont pas de page les décrivant, ne se sont pas même enregistrés par leur pseudonyme, et sont donc uniquement désignés par les IP (*Internet Protocol*), les adresses protocolaires chiffrées des machines depuis lesquelles ils ont effectué leurs sessions. Le but est ainsi d'encourager à participer en gratifiant les nouveaux venus, qui voient leurs corrections immédiatement prises en compte. Sur le Wikipédia francophone, malgré les risques de vandalisme, n'importe quel contributeur, même non enregistré, a la possibilité d'écrire une contribution et même de créer un article. Seuls les deux plus gros Wikipédia ont commencé à instaurer un contrôle éditorial *a priori* : l'anglophone, en 2005, est marqué par l'affaire John Seigenthaler : ce journaliste américain victime de diffamation avait médiatisé sa plainte contre un article où un contributeur anonyme l'avait présenté comme directement impliqué dans l'assassinat du président John Kennedy puis de son frère Robert Kennedy. Or la particularité de ce « vandalisme » était qu'il resta très longtemps sans qu'aucun contributeur ne l'ait relevé<sup>15</sup>. Tirant la conséquence de ce litige, la version anglaise de Wikipédia décida en décembre 2005 que seuls les contributeurs enregistrés pourraient

---

<sup>12</sup> « Nouveaux venus, ayez du culot. » L'adage est une formule centrale dans la grande majorité des communautés dites open source, bien repérée par Paul B. Laat dans un article synthétique sur les sentiments de confiance générés dans ces collectifs, « How can contributors to open-source communities be trusted? On the assumption, inference, and substitution of trust », *Ethics of information and technology*, 2010, n° 12, p. 327-341.

<sup>13</sup> Sur la centration des communautés en ligne sur la serendipité et le paradigme exploratoire, cf. Nicolas Auray, « Folksonomy : a New Way to Serendipity », *Communications & Strategies*, 2007, n° 65, p. 67-91.

<sup>14</sup> Cf. Christophe Aguiton et Dominique Cardon, « The strength of weak cooperation : an attempt to understand the meaning of Web 2.0 », *Communications & Strategies*, 2007, n° 65, p. 51-65.

<sup>15</sup> S'estimant offensé et refusant une action judiciaire incertaine à cause des lois américaines sur la protection de la vie privée qui permettent difficilement de connaître l'identité des clients des fournisseurs d'accès à internet, John Seigenthaler a préféré écrire sur *USA Today* que « Wikipédia est un outil de recherche défectueux et irresponsable » (« Wikipedia is a flawed and irresponsible research tool »). Cf. John Seigenthaler, « A false Wikipedia "biography" », *USA Today*, le 29 novembre 2005.

désormais créer de nouveaux articles. L'enregistrement se limite cependant à la création d'un nom d'utilisateur, qui peut être un pseudonyme, et d'un mot de passe. La mention de l'email est conseillée (ne serait-ce que pour récupérer son mot de passe en cas d'oubli), mais pas obligatoire. En revanche, pas besoin de s'enregistrer pour corriger ou compléter des articles déjà en ligne. De même, depuis septembre 2007, le Wikipédia allemand ne rend visible instantanément que les corrections faites par des utilisateurs « de confiance ». Pour gagner ce statut de confiance, les utilisateurs doivent avoir produit au moins 30 corrections en 30 jours. Les utilisateurs néophytes doivent attendre la validation d'un éditeur de confiance pour que leurs corrections soient prises en compte.

### *L'absence de crainte contre les infiltrations sournoises*

La séparation de l'autorité épistémique et de l'autorité sociale est pratiquée en un sens plus radical que la tolérance à l'absence d'identité : elle s'accompagne d'une acceptation des contributeurs malintentionnés, rémunérés par une organisation ou ayant des objectifs de manipulation stratégique. Étudiants farceurs, gamineries de collégiens, obsédés sexuels, grossièretés de fans dérangés, frappent indifféremment n'importe quel article de l'encyclopédie Wikipédia, comme lorsqu'un internaute vient modifier « Charmide est l'oncle de [Platon] » en « Charmide est l'oncle de [Picsou] ». La tromperie est là temporaire, et vise plus à susciter le rire qu'à manipuler.

La faible interconnaissance génère aussi une forme plus nocive de dégradation : le vandalisme sournois. Celui-ci repose sur des manipulations de l'information qui cherchent à *rester invisibles*. Ainsi, les collectifs en ligne sont souvent investis par des organisations réelles qui détournent l'opinion générale pour dénigrer un concurrent, faire l'apologie de leurs produits, ou se faire de la promotion sur internet. En 2007, un jeune étudiant en neurobiologie de vingt-cinq ans, Virgil Griffith<sup>16</sup>, a par exemple sur Wikipédia mis au point des mouchards pour dresser une liste systématique des manipulations les plus importantes : falsification discrète de mots, changement d'adjectifs... Elles proviennent aussi bien d'entreprises que de collectivités (mairies, États), d'organisations politiques et de sectes. Les « faux avis de consommateurs » regroupent ainsi l'ensemble des pratiques dites, en droit de la concurrence, « déloyales », et qui relèvent soit du baronnage (avis ou commentaires positifs postés par l'entreprise ou un prestataire, sous l'apparence d'un client satisfait) soit du dénigrement des concurrents (avis ou commentaires négatifs dont l'objectif est de nuire à l'image et à la réputation des produits ou services d'autres entreprises concurrentes). En France notamment, les sites internet spécialisés dans la restauration, le tourisme et l'hôtellerie, tels que TripAdvisor, Ciao, Cityvox ou Restoaparis, etc., sont particulièrement concernés car les avis des internautes conditionnent les prises de décision des consommateurs.

La frontière entre la manipulation stratégique et l'apport de contenu augmentant l'intérêt général est très difficile à tracer. Ainsi, un principe d'écriture dans la plupart des communautés en ligne est leur ouverture aux points de vue personnels des membres. Cela est lié au fait que la qualité des œuvres est exprimée différemment par chacun. Même Wikipédia, malgré son principe encyclopédique de « neutralité » de point de vue, définit l'objectivité de manière relativiste, comme « la somme des points de vue pertinents sur un sujet » ; il est valorisé d'introduire un point de vue subjectif, à la seule condition qu'il soit rapporté à son énonciateur et présenté de manière impartiale. Le fondement épistémologique de cette démarche repose sur une conception subjectiviste de la vérité : des faits scientifiques stabilisés sont exposés comme des énoncés renvoyant à un point de vue, tandis que des thèses controversées sont citées comme pertinentes sur un sujet. « Ce que les gens croient, voilà le fait objectif », dit le fondateur de Wikipedia Jimbo Wales<sup>17</sup>. Comme il est légitime d'abonder le corpus en y représentant un point de vue personnel, la lutte contre le vandalisme sournois est donc difficile à mener.

Est-il d'ailleurs réellement prouvé que la lutte contre les manipulations stratégiques apporterait un gain de qualité aux forums ? Certaines modélisations des forums d'échange d'avis sur les biens

---

<sup>16</sup> Cf. le site virgil.gr.

<sup>17</sup> Cf. Jimmy Wales, « Neutral point of view », contribution sur l'article de la branche anglophone de Wikipédia, 2001, disponible en ligne sur web.archive.org .

culturels<sup>18</sup> montrent en effet qu'une éradication des manipulations stratégiques, par la suppression des incitations à contribuer qu'elle entraînerait, pourrait avoir un impact globalement négatif sur le site, en tarissant le nombre d'informations produites. L'enjeu est important, car le vandalisme sournois constitue une pollution du corpus qui peut créer la méfiance des lecteurs. Ils peuvent s'effrayer d'un noyautage possible de la communauté en ligne par une minorité activiste et invisible. La forme spontanée de la crainte est ainsi, en ligne, celle de l'infiltration stratégique. De nombreux collectifs en ligne ont eu à s'en défendre ou à s'en défaire<sup>19</sup>.

### *Le calcul et la détection automatique des « éditeurs de confiance »*

Le travail éditorial repose sur des dispositifs, en grande partie automatisés, d'extraction, à partir du corpus en ligne, de versions *plus stables*. Sur la base d'indices de confiance objectifs, comme par exemple le fait, pour un bout de code informatique, de ne pas s'être vu attribuer un *bug* bloquant de gravité critique durant une période de plus de deux semaines (comme c'est le cas pour définir l'éligibilité d'un paquet à la version stable dans la distribution Debian de systèmes d'exploitations libres), un contenu est automatiquement présélectionné pour être intégré dans une version dite « de confiance ». Une caractéristique de la gouvernance des frontières marquée par le choix de l'ouverture est l'utilisation d'outils et de processus automatiques pour établir, sur la base d'algorithmes, cette sélection et ce nettoyage. Sur Wikipédia, des robots utilisés par des correcteurs bénévoles vérifient la présence de mots offensants dans les modifications récentes, chacun étant accompagné d'un score. Si le total dépasse un certain seuil, le robot révoque la modification. Un problème majeur est la limitation des pouvoirs donnés à ces robots automatisés. Par exemple, une règle centrale a été édictée, la « 1RR » (*one revert rule*, règle d'une seule révocation) pour « dresser » les robots de Wikipédia afin qu'ils ne révoquent pas la même IP sur le même article deux fois de suite. Cette limite inscrite dans le codage des robots vise à donner le pouvoir en dernière instance aux humains qui les utilisent.

Le souci des communautés en ligne est de construire des outils qui permettent de calculer la confiance. Sur Wikipédia, cette opération s'est mise en place au moment où il s'est agi de définir les articles les plus fiables du corpus. Une première proposition fut de sélectionner les contenus de ceux dont les modifications auraient subi le moins de corrections. Cependant, elle fut rejetée car l'encyclopédie est soumise à de constantes modifications de forme qui ne disqualifient pas forcément la pertinence d'un ajout. Wikipédia a alors choisi de miser sur un seuil minimum de travail fourni pour définir « l'éditeur de confiance » : le pari est de considérer qu'on peut prétendre à ce statut à partir de 30 éditions en 30 jours.

## **2- Systèmes de réputation et tyrannies de la visibilité**

Une seconde particularité marque les collectifs en ligne : l'originalité des mécanismes de construction de l'autorité, marqués par le découplage de l'autorité épistémique et de l'autorité sociale. Si l'on entend par « autorité sociale » les représentations sociales et les statuts institutionnels des personnes, les communautés en ligne sont très vigilantes à séparer les critères de construction du statut interne de ces paramètres jugés exogènes, de sorte qu'ils n'influencent pas la façon dont seront évaluées les contributions. Il en découle des mécanismes originaux de construction de l'autorité, marqués par l'invention d'indicateurs statistiques qui « calculent » la

---

<sup>18</sup> Cf. Chrysanthos Dellarocas, Ming Fan et Charles A. Wood, « Self-interest, reciprocity and participation in online reputation systems », *MIT Sloan working paper*, 2004, p. 4500-4504.

<sup>19</sup> Ainsi, parce que Usenet avait des règles très souples de contrôle du corps électoral au moment de ses votes pour la création de nouveaux forums, la communauté fut souvent l'objet de bourrages d'urnes (comme entre septembre 1997 et mars 1998, lors de l'affaire de la création du forum de discussion sur les sectes, fr.soc.sectes, où des prosélytes de la secte Moon diffusèrent un appel massif à aller voter « non »).

réputation. « Karma » dans Slashdot<sup>20</sup>, « étoiles filantes » d'eBay, « hall of fame » (panthéon) des contributeurs sur les forums... Les individus qui contribuent dans les collectifs en ligne se voient attribuer des réputations. Contrairement aux modalités classiques d'organisation du travail dans lesquelles les rôles individuels sont bornés par des limites statutaires rigides prédéfinies, les collectifs en ligne font varier les statuts de leurs membres au gré de leur activité contributive interne, et suscitent un individualisme démonstratif en organisant une course à la réputation individuelle. La publicité donnée à ces échelles de prestige, souvent rappelées sur la page d'accueil du site ou du membre, sert aux collectifs en ligne à stimuler les contributions<sup>21</sup>. Aux membres sont associés des icônes et des scores. Un grand nombre d'étoiles sur eBay ou un bon karma sur Slashdot font remonter la visibilité de leurs détenteurs ou de ce qu'ils disent dans les communautés en ligne. On trouve aussi des « rangs » qui donnent accès à des pouvoirs différents : dans des projets de logiciel libre (comme le système d'exploitation libre FreeBSD) ou dans Wikipédia, un nombre minimum de contributions amène à être éligible pour le rang d'« officier » ou d'« administrateur », qui confère des droits sur la suppression de comptes ou de contenus. La promotion, néanmoins, n'est pas toujours automatique, elle est parfois aussi soumise à un vote complémentaire.

Ce gouvernement de l'activité par l'évaluation publique génère des problèmes de gouvernance particuliers, liés aux risques de biais dans les rapports. Dellarocas a identifié, dans un article de synthèse<sup>22</sup>, les quatre clefs du *design* des mécanismes de réputation: la *sélection* de ceux qui participent à l'évaluation; l'identification du *type de feedback* attendu (une note et/ou un commentaire en texte libre); la détermination du *mode d'agrégation* des avis (la pondération des notes selon les statuts et rangs endogènes); le *type d'information* finalement disséminé à la communauté (moyenne et écart-type ou liste des notes détaillées). Ces mécanismes incluent parfois deux niveaux d'évaluation. Par exemple, sur le site de journalisme participatif Slashdot (300 000 visiteurs par jour), le karma est le fruit d'un calcul savant qui intègre les valeurs des côtes des posts passés et prend en compte l'activité contributive de modération ou de métamodération. Seuls les membres enregistrés et dotés d'un karma positif peuvent modérer : ils disposent pour cela de 5 points de modération toutes les périodes de trois jours, qu'ils peuvent affecter positivement ou négativement sur les contenus postés. Une idée centrale est ainsi qu'il y a un plafonnement des points de modération, pour éviter le noyautage ou le copinage. On ne peut pas non plus modérer un fil dans lequel on a été posteur. Mais, pour éviter les modérations malhonnêtes, il existe aussi des « métamodérateurs ». Pour être éligible à ce statut, il faut être enregistré mais aussi ancien (parmi les 7,5% des comptes les plus anciens) et avoir un karma supérieur à un certain seuil (mobile, selon un choix de la compagnie). Le M2 modère entre 3 et 7 fois par jour, de manière aléatoire, des modérations qui lui sont présentées selon un tirage au sort. La métamodération n'est ainsi pas libre, c'est un peu comme une correction de copie, c'est une activité plus fortement contributive car elle a cette dimension obligée<sup>23</sup>.

### *Les effets réflexifs de la connaissance des scores de réputation*

Face à l'importance de ces signes de statut pour les membres, on a cherché à contrôler la *justesse* de l'échelle de réputation, à maîtriser les effets qu'introduit la connaissance réflexive par les membres de ces communautés des algorithmes de calcul qui produisent la réputation. Cette

---

<sup>20</sup> Site d'actualités anglophones consacré entre autres aux innovations technologiques et aux jeux vidéo.

<sup>21</sup> Cf. Josh Lerner et Jean Tirole, « Some Simple Economics of Open Source », *Journal of Industrial Economics*, 2002, vol. 50, n° 2, p. 197-234.

<sup>22</sup> Dellarocas, C. avec Malone, T., Laubacher, R., « Harnessing Crowds : Mapping the Genome of Collective Intelligence », working paper n°2009-01, Working Center for Collective Intelligence, MIT.

<sup>23</sup> Cf. Lampe, C., Johnston, E., Resnick, P., « Follow the Reader : Filtering Comments on Slashdot », *Proceeding of CHI '07*, pp. 1253-1262.

connaissance réflexive aboutit à des biais qu'il faut sans cesse corriger. D'autre part, elle vise à contrôler et sanctionner les effets de la réputation sur le comportement des membres.

Du fait de la prise de conscience du rôle un peu tyrannique de la réputation, les communautés en ligne peuvent facilement susciter un manque de confiance en soi. Des membres obsédés par les conséquences de leurs comportements sur leur aura virtuelle peuvent s'inhiber au point de refuser d'annoncer ce qu'ils font ou d'en discuter. Ainsi, Yutaka Yamauchi, Makoto Yokozawa, Takeshi Shinohara et Toru Ishida remarquent que, sur les gros projets de développement logiciels, les *patterns* (schémas) interactionnels sont biaisés vers l'action<sup>24</sup>. Les tâches effectuées ne sont pas précédées d'une déclaration préalable. Ce « biais vers l'action » s'explique par son aspect protecteur pour la réputation. Ceux qui n'ont pas grande confiance dans leurs savoir-faire voient se lever les barrières à leur engagement, qui auraient été fortes s'ils avaient été obligés de déclarer publiquement, en arène, leur occupation.

Mais certains effets vont au-delà de l'inhibition, et sont dus au détournement des algorithmes de calcul. Les membres des communautés anticipent les conséquences de chaque pratique sur leur score de réputation future. Le site de ventes entre particuliers eBay offre un bon exemple de « biais » lié à la connaissance par les acteurs des mécanismes de calcul de la réputation. Chaque membre d'eBay possède un profil d'évaluation construit à partir des évaluations laissées par ses interlocuteurs. Les acheteurs et les vendeurs peuvent s'évaluer réciproquement en laissant une évaluation pour chaque transaction. C'est l'acheteur qui commence à mettre son évaluation, après qu'il a reçu le bien, et le vendeur ensuite met la sienne. Les évaluations sont non modifiables et indélébiles (seuls les commentaires, d'un commun accord, peuvent être supprimés). De ce fait, de nombreux acheteurs préfèrent ne pas laisser d'évaluation négative ou neutre lorsqu'ils ont eu une mauvaise expérience, par crainte de recevoir en retour une évaluation négative du vendeur. Les évaluations négatives de représailles sont ainsi devenues monnaie courante. Dès lors, il apparaît un « biais gentil »<sup>25</sup> : le nombre d'évaluations négatives ou neutres laissées sur le site est inférieur au nombre de mauvaises expériences signalées<sup>26</sup>. La connaissance réflexive du calcul de l'évaluation aboutit à une désinformation systématique. La gouvernance des statuts vise à corriger systématiquement ces effets de réflexivité en ajustant en permanence le système pour qu'il soit fidèle aux actes pratiqués. Ainsi, depuis mai 2008, eBay a réformé son système d'évaluation : les vendeurs ne peuvent plus laisser d'évaluations négatives ou neutres aux acheteurs.

**Le risque de surveillance généralisée** A cette inhibition due à la soumission à la visibilité s'ajoute la norme que l'on cherche à plaire, à être vu, pour attirer et retenir l'attention. Le pouvoir aujourd'hui, selon la belle formule de Ganascia<sup>27</sup>, ça n'est plus l'aigle, c'est le paon. L'aigle, c'était celui dont le regard dominait, parce qu'il possédait les informations qui le rendaient plus efficace. Le paon, dont les yeux transportés sur les plumes ne sont plus faits pour voir mais pour être vus, c'est celui qui, parce qu'il est regardé, prend l'ascendant. il y a une inversion du lien entre pouvoir

---

<sup>24</sup> Utilisant les diagrammes de transition fondés sur le codage des actes de langage dans les messages du forum général de discussion, ils montrent ainsi que la probabilité de transition de *report* (rapport) à *question* (question) est nettement supérieure à celle de la transition inverse (de *question* à *report*), qui est négligeable. Les discussions ont lieu après le rapport sur une action, elles ne lui sont pas préalables. La catégorie ayant la plus forte fréquence à l'initiation d'un fil est *report* (32,3 %) devant *question* (22,6 %). De plus, seuls 14,4 % des « rapports » aboutissent à une « question » (donc à une discussion), alors que 28,9 % n'ont pas de relance critique. Cf. Yutaka Yamauchi et al., « Collaboration with Lean Media : How Open-Source Software Succeeds », in *ACM conference on Computer supported cooperative work*, ACM Press, 2000, p. 329-338.

<sup>25</sup> Cf. Chrysanthos Dellarocas, Ming Fan et Charles A. Wood, *op. cit.*

<sup>26</sup> Comme le note un utilisateur du site : « Mi-août, je reçois un mail de vendeur qui ne donne pas suite. J'essaie de négocier, rien à faire, j'ouvre un litige sur eBay. Je place logiquement une note négative au vendeur. En représailles, j'imagine, l'acheteur me place une note négative ! Le comble ! Ma question est donc la suivante : si un acheteur/vendeur victime d'un acheteur/vendeur indélicat ne peut prendre le risque de placer une note négative au risque de s'en prendre une négative en retour, quelle est la valeur d'une évaluation ? » (forum sur le système d'évaluation de eBay, le 5 mars 2007).

<sup>27</sup> Ganascia, G., 2009, *Voir et pouvoir : qui nous surveille ?*, éditions du Pommier.

et regard. Séduire, attirer l'attention, est tellement au centre des activités sur les communautés Harry Potter d'échange et de notation de fictions écrites par les amateurs autour du roman canonique, que l'on a simplifié les normes sociales autour de l'unique préconisation suivante : « R&R » : *Please Read and Review my story*<sup>28</sup>. Une telle soif de reconnaissance peut notamment inhiber l'innovation, en engendrant des comportements moutonniers.

Cette obsession de la visibilité peut en outre faire basculer les collectifs en ligne dans des lieux de surveillance participative généralisée et permanente, marqués par la *délation*, ou dénonciation anonyme, de tout comportement suspect ou bizarre. Déjà, aux USA, l'essor d'Internet dans les petites villes a été marqué par le développement de forums communaux marqués par le commérage et le contrôle social. Ainsi, sur le site Topix, qui agrège et centralise depuis avril 2007 une multitude de forums de discussion locaux, des centaines de citoyens anonymes jettent des accusations sur leurs voisins ; en octobre 2011, ce site, avec plus de 10 millions de visiteurs par mois, était l'un des 10 sites d'information les plus visités du Web américain. De même, un site s'est créé, Cocu 2.0, pour des dénonciations anonymes d'adultères ; le cocu est prévenu : s'il veut connaître le nom de l'amant ou de la maîtresse, il doit déboursier quelques euros.

Parler des gros collectifs médiatisés par internet conduit à définir une forme de confiance *émergente*, une confiance que l'on pourrait dire « ex post » parce qu'elle repose sur le postulat d'un découplage de l'autorité épistémique et de l'autorité sociale, si l'on entend par « autorité sociale » les façons diffuses dont les représentations associées aux inscriptions sociales et institutionnelles des personnes influencent notre jugement sur elles. Se dégager des assignations biographiques et de la tyrannie exercée par les représentations sociales est certes une vieille antienne d'internet, lieu où « le passé et le vrai nom d'un individu [sont] conçus comme lui appartenant de plein droit »<sup>29</sup>.

Que ce soit pour maîtriser l'ouverture des *frontières* de la communauté, ou pour hiérarchiser la visibilité des contenus en fonction de critères d'autorité sur les personnes, les collectifs en ligne adoptent des méthodes centrées sur des « calculs de confiance ». Nous avons aussi détaillé la mise en place de techniques sophistiquées, destinées à permettre le nettoyage automatique des contenus en vertu de règles souveraines, ou à conjurer les menaces de détournement stratégique liés à la transparence, à la connaissance des règles de construction des indicateurs de statut, qui sont publiques.

Cette construction statistique de la communauté renferme, pliée en elle, une promesse politique : celle de la « sagesse de la foule », d'après l'expression de James Surowiecki<sup>30</sup>, qui est la croyance dans la supériorité des vertus de l'intelligence collective sur celles d'une oligarchie experte ou censitaire. La construction démocratique de la confiance s'accompagne, dans les communautés qui ont été impulsées par des amateurs, d'une construction collective des indicateurs selon un savant mixage entre des sélections par la foule et des votes explicites. L'émergence des dispositifs de confiance s'inscrit en effet dans un contexte concurrentiel d'initiatives achevées et parfois contradictoires, desquelles émergent des « recommandations » selon un processus de compétition naturelle. Sur une communauté comme Wikipédia anglophone, par exemple, il existe en parallèle 42 *policies* (principes fondateurs) et 24 *guidelines* (recommandations)<sup>31</sup>, dont la plupart restent à l'état larvaire, écrites sur le site mais non usitées. La confiance pour l'ouverture et la curiosité se double ainsi d'une confiance démocratique.

Ce modèle de confiance, anobli par les gros collectifs d'internet, s'oppose à une conception « institutionnelle » de la confiance. On peut appeler « institution » toute entité qui *confirme* par le langage un état de réalité, dans une situation de mise en doute de la coïncidence entre le langage et la réalité : l'institution académique « certifie » la qualité d'un académique, l'institution juridique

---

<sup>28</sup> Cf. Shirky, C., 2010, *Cognitive Surplus, . Creativity and Generosity in a Connected Age*, Penguin Books, p.90.

<sup>29</sup> La remarque est d'Erving Goffman dans son livre *Stigmate*, Éditions de Minuit, 1975, p. 82, citant Ramon Adams, *The Old Time Cowboy*, The MacMillan Company, 1961.

<sup>30</sup> Cf. James Surowiecki, *La Sagesse des Foules*, J.C. Lattès, 2008.

<sup>31</sup> Cf. Travis Kriplean et al., « Community, consensus, coercion, control: cs\*w or how policy mediates mass participation », *Proceedings of the 2007 international ACM conference on Supporting group work*, ACM Press, 2007.

« certifie » les lettres de créance d'un vendeur. À la différence de ces mécanismes institutionnels, Wikipédia ne repose pas sur les diplômes ; eBay trouve son identité la plus précise dans le fait qu'il n'assume pas la fonction d'adjudicateur, qu'il ne garantit pas la bonne réalisation de la transaction ou la livraison du produit après l'accord, se contentant de fournir des instruments d'appariement et de médiation pour aider les particuliers à résorber leurs différends. Ainsi s'établit la plus grande spécificité des communautés d'internet : installer une confiance « démocratique », qui s'oppose à la confiance institutionnelle<sup>32</sup>.

---

<sup>32</sup> Les travaux du sociologue Niklas Luhmann sont de peu d'utilité pour comprendre cette forme de confiance démocratique : il oppose la confiance institutionnelle – je fais confiance à La Poste, au yaourt, parce que leur force est garantie par des institutions, des tests, des normes de sécurité – et la confiance explicite et décidée – je *prends sur moi de faire confiance à mon voisin pour lui donner une somme d'argent* (ce qu'il appelle *confidence* par opposition à *trust*). Par cette dichotomie, Luhmann ne peut voir l'hybridation qui s'introduit dans la confiance démocratique entre le fait de décider de s'ouvrir aux inconnus (« confidence ») et la force démocratique du nombre (qui serait du « trust », mais dans un sens non institutionnel).