

# Internet et la brutalisation du débat public

*par Romain Badouard*

---

**Internet nous rend-il plus agressifs, ou plus tolérants à l'égard de l'agressivité, dans nos discussions politiques du quotidien ? Romain Badouard dresse une cartographie la violence des débats en ligne, de ses usages et de ses effets.**

---

Le 3 juillet dernier, le tribunal correctionnel de Paris a rendu une décision inédite : pour la première fois en France, deux internautes anonymes ont été condamnés pour cyberharcèlement à 6 mois de prison avec sursis et 2000€ d'amende, après avoir proféré des menaces de mort à l'encontre de la journaliste Nadia Daam. À la suite d'une chronique sur Europe 1 où elle prenait la défense de deux militants féministes eux-mêmes victimes de cyberharcèlement, la journaliste avait fait l'objet d'une campagne particulièrement violente sur des forums de discussion et sur les réseaux sociaux (menaces de viol, appels au meurtre, divulgation de son adresse personnelle, diffusion des photographies de sa fille accompagnées de l'adresse de son école). Au procès, le fossé séparant la violence des propos tenus et les justifications livrées par leurs auteurs ont marqué les journalistes présents sur place : pour ces jeunes adultes, il s'agissait d'humour et de provocations davantage que de réelles menaces. Ils qualifiaient ainsi leurs propos d'échanges « trollesques » entre membres d'une communauté qui partagent les mêmes codes de communication, et où l'ironie est un registre d'expression particulièrement courant. Si cette décision de justice marque une nouvelle étape dans la lutte contre le cyberharcèlement, elle rouvre un vieux dossier pour les chercheurs en sciences sociales qui s'intéressent aux pratiques participatives sur internet : celui de la « qualité » du débat public en ligne.

Internet nous rend-il plus agressifs, ou plus tolérants à l'égard de l'agressivité, dans nos discussions politiques du quotidien ? Pour certains observateurs, il semble indéniable qu'internet, et les réseaux sociaux en particulier, favorisent la « violence expressive », engendrent un « ensauvagement » des relations sociales ou accompagnent une « démocratisation de la méchanceté ». De notre côté, nous préférons aborder le sujet en mobilisant le concept historiographique de « brutalisation », qui désigne initialement un processus d'amplification de la violence politique dans l'entre-deux-guerres, hérité de l'expérience de la Première Guerre Mondiale. Dans le contexte qui nous intéresse ici, toute connotation guerrière ou toute analogie avec le climat politique de l'époque est à proscrire : si nous utilisons le terme de « brutalisation »<sup>1</sup>, c'est pour nous intéresser, comme les historiens qui ont mobilisé ce concept dans leurs travaux, au double processus de banalisation et de légitimation de la violence dans le débat public, qui semble aujourd'hui à l'œuvre dans différents espaces de discussion en ligne.

Les causes de la banalisation de la violence dans les conversations politiques du quotidien sont plus complexes qu'il n'y paraît au premier abord et dépassent largement la seule question de l'anonymat des internautes. Celle-ci présente des racines culturelles, tant les pratiques d'échange propres au web disposent de leurs propres codes, mais aussi sociales, quand l'agressivité en ligne se trouve des justifications morales. Notre objectif est également de montrer que la violence peut constituer une ressource stratégique dans le cadre de controverses autour de sujets de société : faire taire un adversaire en l'intimidant permet alors d'occuper l'espace du débat et d'assurer une visibilité optimale aux arguments que l'on défend. Les plateformes de réseaux sociaux ne sont par ailleurs pas exemptes de responsabilités dans ce dossier, dans la mesure où leur *design* comme leurs modèles économiques favorisent la propagation de contenus virulents, voire haineux. Les pouvoirs publics, de leur côté, incitent les plateformes à s'investir davantage dans la régulation des contenus qu'ils hébergent, laissant paradoxalement courir le risque d'une privatisation des pouvoirs de censure sur internet.

---

<sup>1</sup> Romain Badouard, *Le désenchantement de l'Internet. Désinformation, rumeur, propagande*, FYP Editions, 2017.

## La qualité du débat en ligne

Au début des années 2000, un courant de recherche à la croisée des sciences politiques et des sciences de la communication s'est formé autour de l'analyse des formes de délibération sur internet. Au moyen de méthodes de codage des conversations, les chercheurs et chercheuses de ce champ ont entrepris de caractériser la culture de débat ayant émergé en ligne. L'enjeu était de comprendre si les conditions de prise de parole offertes par internet permettaient aux individus de s'émanciper d'un certain nombre de carcans sociaux, d'évaluer la dimension constructive des conversations en ligne ou encore d'étudier les registres d'expression utilisés pour parler de politique au quotidien. Ces travaux ont mené à des résultats parfois contradictoires<sup>2</sup> : selon certains, internet permet aux individus qui n'osent pas prendre la parole en public de faire valoir leurs arguments dans des conversations politiques<sup>3</sup>, alors que selon d'autres les dynamiques de discussion dans les forums ne font que renforcer la domination des leaders d'opinion et réduire au silence les avis minoritaires<sup>4</sup>. De la même façon, certaines enquêtes mettent en lumière la dimension constructive des discussions quand celles-ci ont pour objectifs de prendre des décisions affectant une communauté<sup>5</sup>, ou quand celles-ci se déroulent dans des espaces non partisans<sup>6</sup>, alors que d'autres soulignent la fréquence des dynamiques d'enfermement idéologique quand les internautes campent sur leurs positions ou radicalisent leurs argumentaires<sup>7</sup>.

La première conclusion que l'on peut tirer de cette somme d'études hétérogènes est qu'on ne débat pas mieux ou moins bien sur internet qu'en face-à-face : on y débat *autrement*. La seconde est qu'il n'existe pas *une* façon de débattre sur internet, mais que chaque espace d'échange présente ses propres standards de discussion en fonction du public qui s'y retrouve (public homogène ou hétérogène), de la finalité de la discussion (parvenir à une décision commune ou échanger de manière informelle) et de

---

<sup>2</sup> Voir à ce sujet la synthèse produite par Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », *Réseaux*, n°150, 2008, p. 19-50.

<sup>3</sup> Laurence Monnoyer-Smith, « Être créatif sous la contrainte. Une analyse des formes nouvelles de la délibération publique », *Politix*, n°75, 2006, p.75-101.

<sup>4</sup> Fabienne Greffet et Stéphanie Wojcik, art. cit.

<sup>5</sup> Voir par exemple le cas de Wikipédia décrit par Dominique Cardon, Julien Levrel, « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, n° 154, 2009, p. 51-89

<sup>6</sup> Scott Wright, « Les conversations politiques en ligne au quotidien : design, délibération et "tiers espace" », *Questions de communication*, n° 30, 2016, p. 119-134.

<sup>7</sup> Voir à ce sujet la synthèse réalisée par Patrice Flichy, « Internet, un outil de la démocratie ? », *La vie des idées*, 2008.

l'architecture de l'espace de débat (les technologies mobilisées pour échanger engendrent des dynamiques plus ou moins délibératives).

Si la « qualité » d'un débat est relative à l'espace dans lequel il prend place, il en va de même pour la perception de l'agressivité : dans de nombreux espaces d'échange en ligne, l'agressivité est un registre d'expression comme un autre et ne constitue pas un obstacle à la bonne tenue de la discussion. Les travaux de Patricia Rossini ont ainsi montré qu'un commentaire vulgaire ou virulent ne l'empêche pas d'être sourcé ou argumenté et qu'en ce sens, l'agressivité intègre les répertoires d'expression légitimes dans de nombreuses communautés en ligne<sup>8</sup>. La tolérance à l'égard des propos virulents se vérifie particulièrement lorsque la violence expressive est tournée vers l'objet de la discussion plutôt que vers un interlocuteur en particulier. Selon Rossini, les propos tenus par les internautes sont proportionnellement plus agressifs dans les sections commentaires des sites de médias que sur Facebook. L'hypothèse avancée par la chercheuse pour expliquer cette constatation est que les discussions dans les sections commentaires sont davantage orientées vers les questions de société ou les personnalités publiques, alors que sur les réseaux sociaux la dimension interpersonnelle de l'échange y prend plus d'importance. Pour le dire autrement, une communauté qui débat se montrera plus tolérante à l'égard d'un message vulgaire ou virulent si l'agressivité est tournée vers une personnalité publique ou un sujet de société, que si elle s'adresse directement à l'une des personnes qui participe à l'échange.

## **Transgresser les normes**

Un autre cas où la violence des propos tenus fait l'objet d'une certaine tolérance est celui des « trolls ». Le qualificatif de « troll » désigne un internaute plus ou moins malveillant dont l'objectif est de « pourrir » des fils de discussion en générant artificiellement des polémiques. Pour ce faire, il exprime un désaccord systématique avec ce qui se dit, dénigre ses interlocuteurs ou tient des propos absurdes dans le seul but d'irriter les autres participants à la discussion. Si le « troll » fait l'objet d'une relative tolérance, c'est que ses prises de parole s'en prennent au cadre de la discussion

---

<sup>8</sup> Patricia G. C. Rossini, « It's Not Deliberation (But That's OK)! Uncivil Discourse, Intolerance, and Online Political Talk », 67e conférence de l'International Communication Association, 28 mai 2017.

davantage qu'aux participants. Leurs interventions pourraient même avoir des effets positifs sur la qualité des débats. Selon le sociologue Antonio Casilli en effet, un troll fédère contre lui l'ensemble des internautes présents dans un fil de discussion, qui vont ainsi dans leurs réponses prendre la défense du collectif et réaffirmer leur appartenance à la communauté. Le troll augmenterait ainsi le niveau d'attention de ces interlocuteurs, stimulerait leur participation et *in fine*, contribuerait indirectement à l'enrichissement de la conversation<sup>9</sup>.

Les cas du « trolling » et celui de l'agressivité comme registre d'expression, que l'on pourrait réunir sous le terme d'« incivilités », doivent être distingués du cyberharcèlement et des discours de haine, même si la frontière entre les genres est parfois poreuse. Dans le premier cas, la véhémence des prises de parole ne constitue pas une attaque contre un interlocuteur, mais une transgression des normes sociales qui régissent un espace d'échange. Dans le second cas en revanche, les propos ont une portée personnelle ou collective qui visent explicitement à dénigrer un individu ou un groupe. Si ces pratiques sont punies par la loi en vertu des dommages psychologiques qu'elles peuvent occasionner chez leurs victimes, elles ont également des effets néfastes sur le débat public. Parce qu'elles ont pour objectif de faire taire un individu ou de discréditer l'opinion d'un groupe, elles peuvent être considérées comme des atteintes au pluralisme démocratique. Elles engendrent des phénomènes de censures collectives et d'auto-censure qui appauvrissent le débat. En octobre 2016, la Commission Européenne a ainsi rendu publique une enquête sur la haine en ligne réalisée dans les 28 États membres de l'Union. Dans cette enquête, les 3/4 des sondés qui indiquaient suivre des débats sur internet affirmaient avoir été témoins de discours de haines, de menaces ou d'insultes. Pour la moitié d'entre eux, ces expériences les avaient dissuadés de prendre part aux conversations alors même qu'ils le souhaitaient initialement.

## **La justification morale de l'agressivité**

Le cyberharcèlement et les discours de haine peuvent pourtant se trouver une justification morale qui légitime leurs usages sur le web et les réseaux sociaux. La chercheuse américaine Lindsay Blackwell a mené une étude dont les résultats vont

---

<sup>9</sup> Antonio A. Casilli, *Les liaisons numériques. Vers une nouvelle sociabilité ?*, Seuil, 2010.

dans ce sens<sup>10</sup>. L'étude en question visait à caractériser les ressorts de l'acceptation de la violence dans le débat en ligne. Pour ce faire, la chercheuse a réuni trois panels différents à qui, à la manière de *focus groups*, elle a demandé de réagir collectivement à un message qu'elle leur présentait. Le message était un tweet, écrit par une certaine « Amy » s'adressant à une interlocutrice nommée « Sarah » en ces termes : « Tu n'es qu'une conne. Suicide-toi. »



Le premier groupe n'a eu accès qu'au message en lui-même et il lui était demandé d'exprimer un niveau d'acceptabilité des propos tenus. Comme on pouvait s'y attendre, ceux-ci ont considéré le message d'Amy comme une transgression grave, condamnable et injustifiée. Le même tweet a par la suite été présenté à un second panel, en lui précisant que la « Sarah » en question avait volé 100 dollars à un couple de personnes âgées. Pour le troisième groupe, la somme en question montait à 10 000 dollars. Dans le cas de ces deux panels, les réactions ont considérablement varié par rapport aux premières réactions : le message d'« Amy » était toujours perçu comme une transgression condamnable, mais celle-ci était majoritairement considérée comme légitime par les membres des panels. Selon Lindsay Blackwell, cette étude illustre la pratique de justice punitive très présente sur les réseaux sociaux, qui s'apparente à la « loi du talion » que l'on résume généralement par l'expression « œil pour œil, dent pour dent ». Les propos injurieux y sont d'autant plus acceptés qu'ils s'adressent à une personne ayant eu un comportement immoral ou qui sort de la norme, et pour lequel elle doit être sanctionnée par la communauté.

Un cas célèbre de justice punitive en ligne est celui de Justine Sacco. En 2013, la jeune britannique part en voyage en Afrique du Sud. Avant de prendre son avion, elle tweete un message disant « Je pars pour l'Afrique. J'espère que je n'attraperai pas le sida. Je plaisante, je suis blanche ». Le tweet en question, qu'elle justifiera par la suite comme un trait d'humour, est très vite repris et condamné sur Twitter pour sa teneur

---

<sup>10</sup> Lindsay Blackwell, «An Eye for an Eye: When Online Harassment is Perceived to be Justified», 67<sup>e</sup> conférence de l'International Communication Association, 28 mai 2017.

raciste. Lorsque Sacco atterrit au Cap, son tweet a occasionné des milliers de réponses : des condamnations, des parodies, des insultes et des appels au meurtre. Le temps du trajet en avion, elle a perdu son emploi (son employeur est directement pris à partie sur Twitter) et une bonne partie de ses relations personnelles et familiales lui tournent le dos. Entre le 20 décembre et la fin de l'année 2013, son nom est recherché 1,2 million de fois sur Google.

## Le lynchage collectif comme distraction

Le cas de Justine Sacco est l'une des histoires qui inspirera au journaliste Jon Ronson son ouvrage *So you've been publicly shamed* qui traite de ces nouvelles formes d'humiliation collective sur les réseaux sociaux. Selon l'auteur, ces lynchages s'apparentent à une nouvelle forme de contrôle social, voire une résurgence des tribunaux populaires ou des jeux du cirque. Sans aller jusque-là, il semble indéniable que ces formes de condamnation collectives s'inscrivent dans une histoire. Selon le chercheur François Jost, elles peuvent être considérées comme les héritières du « *hate watching* », cette pratique qui vise à regarder des émissions tout en en détestant le contenu, dans le seul but de les critiquer et de tourner en ridicule ses protagonistes. Les réseaux sociaux prendraient ainsi le relais de la télé-réalité du début des années 2000, en amplifiant la dimension collective de la moquerie et par effet d'entraînement, sa virulence<sup>11</sup>.

Le lynchage en ligne comme pratique sociale, voire comme distraction, soulève la question de sa réception. Si l'ironie et le second degré sont des registres d'expression particulièrement mobilisés sur les réseaux sociaux, tous les internautes ne disposent pas de la même grille d'interprétation des messages violents. Les chercheuses Amanda Lenhart et Michele Ybarra ont par exemple mené une étude portant sur la perception du cyberharcèlement chez les adolescents aux États-Unis<sup>12</sup>. Elle révèle que si les garçons sont autant victimes d'agressions verbales que les filles (dans cette étude, un adolescent sur deux se dit avoir été la cible d'attaques en ligne), le discernement entre ce qui relève de l'agression et de l'humour varie considérablement selon les sexes : les jeunes femmes, davantage sujettes à des insultes concernant leur physique, expriment

---

<sup>11</sup> François Jost, *La méchanceté en actes à l'ère numérique*, CNRS Editions, 2018.

<sup>12</sup> Amanda Lenhart, Michele Ybarra, « But It Was Just A Joke ! Online Harassment and the Differential Understanding of What Constitutes Abuse Between American Men and Women », 67<sup>e</sup> conférence de l'International Communication Association, 28 mai 2017.

un seuil de tolérance inférieur à la violence expressive, alors que pour les jeunes hommes, un propos est considéré comme « admissible » tant qu'il ne va pas jusqu'à la menace directe de violence physique.

Que des insultes, des propos virulents ou des propos haineux soient formulés au premier ou au second degré, il n'en demeure pas moins que leur prolifération engendre irrémédiablement leur banalisation. Comme le veut la « loi de Poe », en ligne, les extrémismes « sincères » et « satiriques » sont difficilement différenciables et constituent en fin de compte les deux facettes d'un même processus de normalisation de l'agressivité. La sociologue Catherine Blaya a récemment conduit une enquête portant sur l'exposition des adolescents aux discours haineux sur internet, dont les conclusions vont dans ce sens. À partir de 1 093 questionnaires et entretiens réalisés auprès de collégiens et lycéens âgés de 11 à 18 ans, elle a estimé que 35,2 % d'entre eux avaient été confrontés à des messages racistes, antisémites, misogynes ou homophobes sur internet, et que 15% d'entre eux en avaient été directement victimes. L'enquête en question relevait que le sexe ou l'origine sociale n'avaient pas d'incidence sur le fait de devenir producteur de haine. Elle montrait par ailleurs qu'une part importante des auteurs de ces injures en étaient eux-mêmes victimes. Les discours haineux semblaient ainsi engendrer un cercle vicieux : puisque les internautes visés en étaient victimes dans l'indifférence générale, ils les intégraient comme des formes d'expression légitimes et les reproduisaient à leur tour<sup>13</sup>.

## L'anonymat en question

Lorsqu'il s'agit d'aborder la question du cyberharcèlement ou des discours de haine en ligne, l'anonymat des internautes est souvent désigné comme la cause principale de leur prolifération, dans la mesure où il favoriserait une forme de déresponsabilisation des individus. Protégés derrière un pseudonyme, ils trouveraient sur internet un exutoire idéal pour manifester leurs frustrations sans avoir à en payer les conséquences. Ces accusations oublient souvent que l'anonymat constitue l'un des piliers de la culture de débat en ligne, et qu'il est une condition *sine qua none* à la libération de la parole permise par internet. Dans notre vie quotidienne en effet, nos prises de parole sont en grande partie conditionnées par un ensemble de marqueurs

---

<sup>13</sup> Catherine Blaya, « Exposition et incitation à la haine sur internet », Conférence Les carrières de la terreur », CNRS, 8 juin 2016.



sociaux. On sait que lors de réunions publiques par exemple, les femmes se manifestent moins que les hommes et qu'elles sont davantage sujettes à se faire interrompre. De la même façon, les individus qui disposent d'un faible niveau d'éducation osent moins prendre la parole dans une conversation lorsqu'elles perçoivent leurs interlocuteurs comme davantage cultivés qu'elles. L'émergence d'un internet grand public au milieu des années 1990 a porté en elle la promesse de pouvoir s'affranchir de ces carcans. Sur un forum ou un réseau social, personne ne perçoit l'âge, le sexe ou la couleur de peau de son interlocuteur. On peut y aborder les sujets que l'on n'ose pas aborder avec ses proches en face-à-face, on peut tester la pertinence d'un argument ou la popularité d'une opinion.

Il semblerait par ailleurs que l'anonymat ne constitue pas réellement un élément provocateur de l'agressivité dans le débat. Des chercheurs de l'université de Zurich ont par exemple étudié récemment plus de 500 000 commentaires sur une plateforme de pétitions allemande. Leur but était d'observer la manière dont les signataires débattaient de sujets sensibles en s'intéressant notamment aux registres d'expression auxquels ils avaient recours. Leur conclusion est qu'un individu qui tient des propos agressifs sous sa propre identité sera jugé comme plus « crédible » par ses interlocuteurs parce qu'il assume ses opinions. La radicalité produit une forme de reconnaissance particulièrement recherchée lorsque le sujet du débat est vague ou charrie un nombre important d'incertitudes. Les chercheurs suisses soulignaient ainsi que plus la thématique d'un échange était controversée, plus les propos violents exprimés à découvert étaient nombreux.

## **La violence comme ressource stratégique**

On retrouve cette idée de l'agressivité comme ressource stratégique dans les propos tenus par les internautes ayant harcelé Nadia Daam. Lors de leur procès, lorsque le juge demande à l'un des deux accusés pourquoi il a tenu de tels propos sur un forum, celui-ci répond : « je ne vais pas dire que c'était pour gagner des points sur le forum ou des points internet mais... c'est presque ça ». Après la condamnation des deux internautes, un troisième individu postera sur le même forum un nouvel appel au meurtre. Il sera à son tour jugé en comparution immédiate. À la barre, il plaidera une « blague » pour divertir la communauté. Comme le rapporte la journaliste du Monde présente à l'audience : « il reconnaît seulement des propos « aguicheurs », proférés lors d'un « pic de connexion ». (...) Dans le flot de messages publiés sur la journaliste ce jour-

là, le prévenu cherche alors à se distinguer de la multitude : ‘Je me suis dit que j’allais faire mieux que les autres’. » Les propos virulents constituent ici un moyen de gagner en reconnaissance sur le forum, d’apparaître comme une personne ressource, voire un « leader d’opinion ».

Les usages stratégiques de la violence expressives dépassent pour autant les seules logiques de reconnaissance interpersonnelle et présentent des causes davantage structurelles. Le débat public en ligne s’apparente bien souvent à des batailles de visibilité quand les sujets abordés sont controversés : les partisans d’une cause vont chercher à rendre leurs argumentaires le plus visible possible aux yeux des internautes, par exemple en améliorant le référencement de leurs sites sur les moteurs de recherche (pour s’assurer qu’ils disposent d’un bon classement et arrivent dans les premiers résultats) ou en ayant recours à des hashtags sur les réseaux sociaux afin de faire remonter les fils de discussion dans les « sujets tendances » (*trending topics*) sur lesquels tomberont automatiquement les internautes en s’y connectant. Sur Twitter, le recours aux hashtags dans les conversations relève même parfois d’une « politique affective » pour reprendre l’expression de Maxime Cervulle et Fred Pailler : des camps opposés sur une même thématique vont se doter de leurs propres hashtags pour s’attribuer des espaces de débat militants (par exemple #mariagepourtous et #manifpourtous lors des débats sur la légalisation du mariage homosexuel), et rassembler des sympathisants autour de visions communes. Pour autant, la plasticité des hashtags couplée au système d’adressage permis par l’arobase sur Twitter (qui permet de mentionner directement quelqu’un dans son tweet et de le lui faire savoir) produisent des ponts entre ces communautés idéologiques et génèrent des polémiques et du conflit sur le réseau social<sup>14</sup>.

Tout l’enjeu, dans ces conflits, est moins de convaincre des opposants dont les positions sont bien établies que de s’adresser à la majorité silencieuse qui consulte les débats sans jamais y prendre part. Si l’architecture des réseaux sociaux est conçue de manière à inciter les internautes à prendre la parole ou à réagir, il semblerait que la plus grande partie d’entre eux utilisent ces plateformes à la manière de navigateurs, pour accéder à des contenus sans jamais participer aux conversations, « liker » des messages où partager des *posts*. Des camps en opposition vont ainsi chercher à diminuer la visibilité des arguments adverses, soit en augmentant la visibilité de leurs propres arguments, soit en cherchant à disqualifier leurs opposants, afin de les délégitimer aux yeux du public « neutre » qui assiste à la joute. Dans ce contexte, faire

---

<sup>14</sup> Arnaud Mercier, « Twitter, espace politique, espace polémique. L’exemple des tweet-campagnes municipales en France (janvier-mars 2014) », *Les Cahiers du numérique*, vol. 11, 2014, p. 145-168.

taire ses adversaires en les intimidant peut relever d'une stratégie collective. Le recours au cyberharcèlement militant tend ainsi à devenir un mode d'action collective à part entière sur les réseaux sociaux. Cette pratique, historiquement liée aux réseaux d'extrême droite parfois qualifiés de « fachosphère », se serait étendue aux militants des causes progressistes. Jonathan Chait, journaliste au *New York Times*, a dénoncé dans une tribune remarquée une nouvelle culture du politiquement correct au sein de la gauche radicale américaine qui viserait à disqualifier par l'humiliation les points de vue considérés comme « illégitimes ». Ces pratiques auraient selon le journaliste pour conséquence de limiter la liberté d'expression d'individus qui craignent d'être perçus et affichés comme « rétrogrades » sur les réseaux sociaux. « Le débat démocratique repose sur l'idée de convaincre les gens d'être d'accord avec vous », conclut Chait dans cette tribune, « pas de leur faire redouter de ne pas être d'accord. »

## La responsabilité des plateformes

Face à cette situation, les plateformes de réseaux sociaux, où se déroule aujourd'hui l'essentiel du débat public en ligne, ont une première responsabilité liée à la conception des espaces de débat qu'ils mettent à disposition de leurs utilisateurs. Le design des outils d'expression a en effet une influence primordiale sur les dynamiques d'échange qu'ils font naître. Sur Twitter par exemple, le recours aux hashtags, l'adressage via l'arobase ou la limite des 140 caractères (aujourd'hui 280) favoriseraient des dynamiques de simplification et de radicalisation des propos, propices à des échanges conflictuels<sup>15</sup>. Les algorithmes qui trient et hiérarchisent les informations sur ces plateformes sont également régulièrement mis en cause. Sur Facebook, la manière dont le EdgeRank organise les posts des contacts des utilisateurs sur leurs fils d'actualité engendrerait la constitution de « bulles informationnelles » où les internautes seraient sans cesse au contact d'informations qui les confortent dans leurs opinions. Plus globalement, la conception des espaces de débat sur les réseaux sociaux est principalement tournée vers l'expression de préférences, par exemple à travers les « likes » sur Facebook ou YouTube. Ceux-ci seraient ainsi relativement peu utilisés comme des dispositifs de prise de parole en tant que tels, et consacraient à l'inverse

---

<sup>15</sup> Julien Longhi, « Essai de caractérisation du tweet politique », *L'Information grammaticale*, n° 136, 2013, p. 25-32.

une posture réactive chez les internautes, souvent exprimée en termes binaires (j'aime / j'aime pas), qui ne facilite pas l'épanouissement d'un débat constructif et apaisé<sup>16</sup>.

La seconde responsabilité des plateformes a trait à leur modèle économique. La plupart des grandes entreprises du web génèrent l'essentiel de leur chiffre d'affaires en vendant l'attention de leurs utilisateurs à des annonceurs : plus les internautes passent de temps à utiliser leurs services, plus ils sont exposés à des publicités, et plus les plateformes génèrent des revenus. Dans ce contexte, les contenus haineux, au même titre que les contenus mensongers, constituent des produits informationnels particulièrement compétitifs. Parce qu'ils génèrent de l'« engagement », c'est-à-dire qu'ils font réagir les internautes et suscitent des visionnages, des clics et des partages, ils contribuent à la croissance économique des plateformes. Ainsi, au-delà d'un simple « laisser-faire », ces dernières sont parfois accusées d'assurer la promotion de contenus douteux. L'algorithme de recommandation de YouTube par exemple, a été récemment mis en cause dans la promotion de vidéos complotistes : les internautes passant du temps à les visionner, elles permettent d'y insérer davantage de « coupures publicitaires », quitte à les recommander à un public qui ne chercherait pas a priori ce type de contenu. Ce qui avait fait polémique dans cette affaire est que cette même logique s'appliquait à la plateforme YouTube Kids, pourtant destinée aux enfants à partir de 4 ans.

## Lutter contre la haine en ligne

La « responsabilisation des plateformes » est ainsi au cœur des politiques publiques de lutte contre la haine en ligne. Historiquement, les géants du web ont toujours adopté une posture de « plombiers », qui gèrent des « tuyaux » mais ne souhaitent pas intervenir dans la régulation des contenus qui y circulent. Ces entreprises ont trainé des pieds, à de multiples reprises par le passé, pour retirer des contenus haineux, déréférencer des sites, supprimer des *posts* et des contenus. Si la logique semble s'être inversée depuis les attentats de 2015 et la diffusion de propagande djihadiste sur les réseaux sociaux, la nouvelle posture des plateformes (intervenir davantage, et plus rapidement) s'accompagne également de risques de délégation de pouvoirs de censure des états vers des entreprises privées. En Allemagne, une loi particulièrement contraignante pour les plateformes de réseaux

---

<sup>16</sup> Voir à ce sujet les travaux d'Olivier Houdé en sciences cognitives.

sociaux est entrée en vigueur en janvier 2018. Elle les oblige notamment à retirer les contenus haineux qui leur sont signalés en moins de 24h sous peine d'amendes pouvant atteindre les 50 millions d'euros. L'application de cette loi fait aujourd'hui polémique outre-Rhin dans la mesure où les plateformes sont maintenant accusées d'excès de zèle : elles censureraient des contenus qui ne relèveraient pas de propos illicites, et ceci en toute opacité.

La lutte contre les discours de haine en ligne est un exercice particulièrement périlleux dans la mesure où elle peut rapidement constituer une atteinte aux libertés fondamentales sur internet. Dans ce contexte, lutter contre les contenus illicites doit également s'accompagner d'une protection de la liberté d'information et d'expression des internautes, notamment en obligeant les plateformes à plus de transparence sur leurs pratiques de censure (communiquer sur les contenus retirés et les critères à partir desquels ils sont retirés) et en offrant aux internautes des voies de recours ou de contestation quand leurs contenus sont supprimés. Le récent rapport « Renforcer la lutte contre le racisme et l'antisémitisme sur internet », remis au Premier ministre le 20 septembre dernier, émet des propositions en ce sens. Jusqu'à présent, la lutte contre les propos haineux en ligne repose principalement sur le volontarisme des plateformes, lui-même alimenté par le besoin de protéger leur image de marque. À l'échelle européenne, un code de conduite visant à lutter contre les discours haineux illégaux en ligne a été mis en place en juin 2016. À cette occasion, les principales plateformes ont pris un certain nombre d'engagements, en évitant cependant toute mesure contraignante. La Commission se satisfait pour l'instant de cette situation, qui semble par ailleurs porter ses fruits : d'après une enquête réalisée par ses soins fin 2017, les entreprises signataires retireraient 70% des contenus qui leur sont signalés (contre 59% lors de la précédente enquête, en mai 2017), et dans 81% des cas en moins de 24h.

Une autre piste, qui pourrait s'avérer particulièrement fructueuse mais qui constitue jusqu'à présent le parent pauvre des politiques publiques de régulation des contenus, serait de s'attaquer au marché de la publicité en ligne. En imposant par exemple aux régies publicitaires de rendre publique la liste des sites sur lesquels sont diffusées les publicités des annonceurs, ces derniers seraient incités à refuser que leurs campagnes soient relayées par des sites hébergeant des contenus haineux. De la même façon, les réseaux sociaux pourraient être contraints à davantage de transparence concernant les individus et les organisations qui sponsorisent des contenus douteux via leurs propres régies. Démonétiser les contenus qui propagent des discours de haine pourrait constituer un frein efficace à leur circulation. Ces pistes sont également

abordées dans le rapport mentionné plus haut, reste maintenant à savoir si le législateur s'en saisira.

Enfin, la lutte contre les discours de haine ne concerne pas que les plateformes et les pouvoirs publics, il est l'affaire de tous. Des associations ont ainsi entrepris de mettre en place des stratégies de contre-discours sur les réseaux sociaux, visant à mobiliser les internautes témoins de discours de haine à leur porter la contradiction en intervenant dans le débat. La plateforme Seriously, portée par le think tank Renaissance Numérique, ou le site Répondre aux préjugés, proposent par exemple aux internautes des banques d'arguments et de statistiques à poster en réponse aux propos racistes, antisémites, homophobes ou misogynes. L'ambition est ici de chercher à occuper l'espace du débat pour délégitimer les contenus haineux aux yeux du public qui assiste aux joutes sans y prendre part. Ces initiatives accompagnent également une nouvelle approche de l'éducation aux médias et à l'information à l'école, visant à promouvoir une « éducation au débat » qui, au-delà de l'évaluation de la pertinence des sources d'information, s'intéresse à leur appropriation par les enfants et les adolescents et aux mécanismes de construction d'opinions collectives dans l'échange.

### Pour aller plus loin

- Romain Badouard, *Le désenchantement de l'internet. Désinformation, rumeur, propagande*, FYP Editions, 2017.
- Dominique Cardon, *La démocratie Internet*, Le Seuil, 2010.
- Antonio Casilli, *Les liaisons numériques. Vers une nouvelle sociabilité ?*, Le Seuil, 2010.
- Patrice Flichy, « Internet, un outil de la démocratie ? », *La vie des idées*, 2008  
<https://laviedesidees.fr/Internet-un-outil-de-la-democratie.html>
- Fabienne Greffet, Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxon », *Réseaux*, n°150, 2008, p. 19-50.
- François Jost, *La méchanceté en actes à l'ère numérique*, CNRS Editions, 2018.
- Arnaud Mercier, « L'ensauvagement du web », *The Conversation*, 2018.
- Scott Wright, « Les conversations politiques en ligne au quotidien : design, délibération et "tiers espace" », *Questions de communication*, n° 30, 2016, p. 119-134.

**Ressources en ligne :**

- [Internet Actu](#)
- [Affordance.info](#)

Publié dans [laviedesidees.fr](http://laviedesidees.fr), le 6 novembre 2018.