

Dans le miroir des réseaux sociaux

par Simon Chauchard

Quel rôle les médias sociaux jouent-ils dans la polarisation croissante des sociétés occidentales ? Chris Bail mène l'enquête, et pointe du doigt les utilisateurs eux-mêmes.

Au sujet de : Chris Bail, [Breaking the Social Media Prism: How to Make our Platforms Less Polarizing](#), Princeton University Press (2021), 240 p.

Au cours des dix dernières années, la société américaine a connu une polarisation sans précédent. Les opinions politiques des Républicains et des Démocrates sont devenues de plus en plus distantes. Le phénomène a aussi pris un tour nouveau, plus affectif : les individus de bords opposés ont perdu la capacité de se fréquenter et ont développé des niveaux de détestation jusqu'ici peu imaginables. Des tendances similaires semblent exister dans d'autres démocraties.

Depuis l'élection de Donald J. Trump en 2016, les réseaux sociaux ont souvent été tenus responsables du problème.

La critique des plateformes des géants de la Silicon Valley repose sur trois arguments principaux. Le premier consiste à dire que les médias sociaux n'exposent les individus qu'à des opinions similaires aux leurs. D'après cette idée, les médias sociaux font vivre les utilisateurs dans des bulles d'opinions qui ne peuvent que renforcer leurs opinions existantes, puisque limitant mécaniquement la diversité des sources et des opinions auxquelles ils sont exposés (le phénomène des « chambres d'échos »). Le deuxième argument avancé est que les médias sociaux servent de véhicules à de puissantes campagnes de désinformation qui peuvent, dans le pire des cas, faire perdre ou gagner une élection. La troisième idée aggrave le constat fait sur

les chambres d'échos : selon cet argument, les algorithmes de certaines plateformes contribuent non seulement à la polarisation des individus, mais également à leur radicalisation progressive, en les entraînant vers des idéologies radicales ou des groupes extrémistes qu'ils ne fréquentaient pas précédemment.

Mis ensemble, ces arguments nous proposent une vision résolument catastrophiste de l'influence des réseaux sociaux sur la société : en choisissant pour nous les individus et les contenus avec lesquels nous interagissons, ces plateformes auraient acquis un fort pouvoir de persuasion sur les populations, les entraînant dans une spirale inexorable vers les extrêmes.

Cette vision est aujourd'hui dominante. Des critiques les plus alarmistes aux plateformes elles-mêmes, plus personne ne conteste aujourd'hui le fait que les médias sociaux représentent un risque pour la démocratie. De même, la polarisation grandissante des sociétés occidentales ne fait plus guère l'objet de débats.

Et si les réseaux sociaux n'étaient pas responsables de tout ?

Pour autant, ni les mesures à prendre pour remédier à cette polarisation ni les mécanismes liant médias sociaux et polarisation ne sont clairs.

Dans cet ouvrage, Chris Bail, professeur de sociologie quantitative et de *data science* à l'Université de Duke, s'attache à mieux comprendre les liens entre polarisation et médias sociaux. L'approche se distingue des dizaines d'ouvrages publiés depuis 2016 par une approche résolument empirique. Au lieu de *présupposer* les arguments cités précédemment, Bail s'appuie sur le meilleur de la *data science* américaine pour les *évaluer*. En ce sens, bien qu'écrit pour un large public, le livre est en partie une recension de la production récente des meilleurs spécialistes américains au sujet de l'influence possible des médias sociaux sur l'opinion. Au sein de ces études figurent les propres travaux de l'auteur sur les interactions entre utilisateurs de Twitter, ainsi qu'une série d'enquêtes longitudinales qui permettent à l'auteur d'illustrer ses arguments.

Les travaux mis en avant par Bail ont un point commun. Malgré une limitation commune d'importance (tous sont centrés sur un seul cas, celui des interactions sur

Twitter aux États-Unis), ils remettent en cause la vision selon laquelle les médias sociaux exercent un pouvoir d'influence extrême.

S'appuyant sur d'impressionnantes données, Bail montre que très peu d'individus sont *in fine* entraînés vers des idéologies radicales par leur fréquentation de Twitter, ou de YouTube. Bail et ses collaborateurs montrent également que les campagnes de désinformation russes n'ont probablement pas eu un effet susceptible (par leur magnitude) de changer le résultat d'une élection.

Miroirs déformants et Estime de soi

Mais le véritable cœur du livre est la réfutation appuyée de la thèse selon laquelle les chambres d'échos mènent à la polarisation. L'auteur s'appuie ici sur une expérience remarquable et remarquée, puisque récipiendaire de nombreux prix depuis. Pour évaluer le rôle que les chambres d'échos jouent dans la polarisation, Bail et ses collaborateurs ont recours à une solution aussi inhabituelle qu'ambitieuse : ils modifient artificiellement le contenu Twitter d'un large groupe de républicains et de démocrates sur Twitter. Sans leur révéler le but de l'expérience et sans se référer explicitement aux identités partisans (qui sont mesurées implicitement par le biais de leurs tweets et retweets précédents), Bail incite ces utilisateurs de Twitter à suivre certains comptes pendant une période d'un mois, sans leur expliquer le motif réel de cette démarche.

Or, ces comptes sont contrôlés par l'équipe de recherche et relaient des positions contraires à celles des participants. Le but n'est ici pas d'entraîner les individus vers une position centriste, mais plutôt de modérer leurs sentiments envers les individus du bord opposé. Mais rien de cela n'arrive. Bien qu'ils aient un côté inévitablement artificiel, les résultats remettent en cause le constat dominant sur les chambres d'échos. Tandis que l'ambition était d'extraire les utilisateurs de leurs bulles respectives afin de modérer leurs sentiments, l'expérience radicalise les sujets et rend leurs opinions encore plus extrêmes et plus antagonistes (Bail parle d' « effet de retour de flamme »). Les raisons de cette réaction demeurent quelque peu mystérieuses. Quoi qu'il en soit, cela implique que sortir les individus de leurs chambres d'échos ne constitue guère une solution à la polarisation, et que cela pourrait même l'aggraver. Par extension, il semble que la polarisation en ligne ne soit pas seulement due à ces mêmes chambres d'échos.

Si les chambres d'échos ne sont pas responsables de la polarisation, quelles sont les causes du phénomène ? Les lecteurs pourront être tentés d'invoquer la fuite en avant du parti républicain, les tendances oligarchiques du système américain, ou le pouvoir d'influence de Fox News. Mais Bail ignore totalement ces forces manifestes pour se concentrer sur les motivations intimes et psychologiques des individus. Bien que cet angle mort soit une autre faiblesse de l'ouvrage, le choix permet aussi à Bail de se concentrer sur le rôle des utilisateurs. Car selon Bail, les utilisateurs ont bien une part de responsabilité.

Selon Bail, les médias sociaux mènent à la polarisation en partie parce qu'ils sont le lieu de la présentation de soi, de la cultivation de l'image, de l'estime de soi, et finalement, de la gratification narcissique. D'après Bail, les médias sociaux nous servent avant tout à travailler notre image et à nous présenter au monde, et ainsi à nous sentir membres d'un ou plusieurs groupes. Tel un politicien engageant un institut de sondage pour en savoir plus sur son image avant une élection, nous avons recours aux réseaux sociaux afin d'obtenir une source infinie d'évaluations sur nous-mêmes. Que sont les « likes » et autres « cœurs » que nos contacts cliquent, ou pas, en réponse à nos saillies en ligne ?

Selon Bail, cette caractéristique centrale de Facebook et Twitter nous donne l'impression d'avoir acquis une capacité renforcée à évaluer ce que les autres pensent de nous et au-delà, ce que la société pense. Le problème est que cette impression est largement erronée : le « prisme des médias sociaux » qui donne son titre à l'ouvrage est l'idée que les plateformes nous offrent une vision déformée de la réalité. Sur ce point, Bail se base largement sur de longues interviews d'utilisateurs dont il analyse, en parallèle, le comportement sur Twitter. Puisant dans ces données qualitativement riches, Bail construit un modèle explicatif qui brille autant qu'il pêche par sa simplicité.

Son argument central est que les plateformes exagèrent la place d'un petit groupe d'utilisateurs décrits comme « extrémistes ». La raison est simple : la majorité des utilisateurs – les « modérés », dans la classification binaire de Bail - sont effrayés par la nature des interactions en ligne. Par conséquent, ils contribuent peu, ou pas, de peur de s'attirer les foudres d'utilisateurs plus féroces qu'eux-mêmes. Cette retenue laisse la place aux « extrémistes », qui eux savent se satisfaire des interactions peu civiles que les médias sociaux encouragent, dans leur architecture actuelle. Mises ensemble, ces tendances déforment notre vision de la réalité politique de la société, puisqu'elles faussent la représentativité des opinions existantes dans la société en donnant une place trop grande aux extrémistes. Cette « fausse polarisation », basée sur

une vision déformée de l'opinion publique, n'est pour autant pas neutre, puisqu'elle contribue à générer de la « vraie polarisation » : si nous pensons que les opinions extrémistes auquel le prisme des médias sociaux nous expose sont représentatives de l'autre bord, nos sentiments envers nos adversaires politiques empirent.

Réinventer les Médias sociaux ?

Puisque c'est là que le nerf de la guerre se situe, comment donc effacer les ressorts psychologiques qui mènent les extrémistes à prendre le dessus sur les modérés en ligne ?

La dernière partie du livre s'attache à proposer des mesures dans cette direction. Selon Bail, une grande partie de l'entreprise consiste pour chacun d'entre nous à identifier l'existence de ce prisme déformant, pour mieux le contrer. L'idée peut paraître simpliste, mais a le mérite de redonner le pouvoir aux utilisateurs : si nous devenons conscients du fait que les opinions observées en ligne représentent mal la société, que nos opinions (généralement modérées) sont communes, ou que nous postons essentiellement pour « appartenir », ce prisme peut à terme disparaître. Bail propose une série d'outils cognitifs pour aider les utilisateurs de Twitter en ce sens, comme un outil montrant leur place (en général, bien plus modérée qu'ils ne l'attendent) sur le spectre idéologique en analysant leurs tweets, ou un outil montrant que les points de vue extrémistes sont en réalité rares.

Au-delà de ces incitations, dont on peut douter de l'efficacité sur des utilisateurs par ailleurs accros aux réseaux (la métaphore des drogues est récurrente dans le livre), Bail concède qu'un différent type de média social sera éventuellement nécessaire pour atteindre cet objectif. Il l'imagine et le teste dans une expérience ouvertement utopique décrite à la fin de l'ouvrage. *DiscussIt*, la plateforme conçue par son laboratoire, est conçue comme un lieu où les individus discutent anonymement de sujets variés. Après avoir reçu des pseudonymes ne révélant aucune identité collective, et après avoir été déconnectés de leurs milieux sociaux par le fait d'une anonymisation totale, les utilisateurs sont invités à tester cette nouvelle plateforme. Une fois sur la plateforme, ils sont connectés sans le savoir à des utilisateurs de l'autre bord et débattent d'un sujet qui leur est proposé. Même s'il est difficile d'imaginer comment un entrepreneur aurait intérêt à créer une telle plateforme, l'expérience permet de montrer que les médias sociaux peuvent *en théorie* permettre des interactions plus constructives. Les

résultats sont en effet spectaculaires : la plupart des sujets acquièrent rapidement une opinion plus favorable de l'autre camp et la participation se rééquilibre au profit des « modérés ».

Ces résultats impliquent qu'il est nécessaire de sortir les individus de leur chambre d'écho, mais qu'il faut encore, pour atténuer la polarisation inhérente aux interactions en ligne et faciliter l'échange, les déconnecter de leurs réseaux numériques. Ayant perdu la capacité de s'ériger en champions d'une cause collective, Bail montre que la minorité « extrémiste » évolue vers des comportements moins négatifs, faisant par là même une place aux utilisateurs « modérés ». Poussant l'imagination plus loin, Bail se met à rêver de plateformes qui favoriseraient algorithmiquement les messages qui résonnent parmi des groupes opposés, d'un remplacement des « like » par des compteurs qui identifient de tels contenus, ou d'incitations automatisées à la réflexion avant de poster des messages identifiés par l'intelligence artificielle comme ayant un potentiel de division.

Certaines de ces suggestions paraissent plus judicieuses que d'autres. Certaines paraissent clairement risquées, comme l'anonymisation, dont les effets sur le harcèlement en ligne sont par ailleurs connus ; la plupart sont plus simplement contraires aux intérêts économiques de la Silicon Valley, du moins à court terme, et en cela irréalistes. Mais toutes sont techniquement réalisables, et à même d'être testées dès aujourd'hui pour réinventer les médias sociaux. En ce sens, les travaux de Bail doivent être pris au sérieux, malgré leurs limitations empiriques. Sans clairement montrer la voie à suivre, ils prouvent tout au moins que la polarisation tient de l'architecture actuelle de ces plateformes plutôt que leur existence intrinsèque, et que la régulation pourrait faire beaucoup pour atténuer ces effets pervers.

Publié dans lavedesidees.fr, le 1^{er} septembre 2021.