Il faut faire attention aux fausses nouvelles

Pourquoi les activités d'influence posent-elles un défi aux gouvernements démocratiques et libéraux

Eric Jardine

L'élection présidentielle américaine de 2016 a été marquée par le tumulte. Durant les mois et les semaines qui ont précédé ce mardi 8 novembre, les sites des médias sociaux, tels que Twitter et Facebook, ont été inondés de « fausses nouvelles » (Howard et coll., 2017). Les enquêtes menées suite à l'élection de Donald Trump à titre de quarante-cinquième président des États-Unis ont révélé qu'une influence étrangère majeure avait joué un rôle durant la campagne, essentiellement dans le but de modifier le cours de l'élection. La majorité des doigts se sont alors pointés en direction des coupables présumés : la Fédération de Russie et le régime du président Vladimir Putin (National Intelligence Council, 2017).

Il ne s'agissait en aucun cas de la première utilisation des médias sociaux à des fins d'activités d'influence. En effet, quelques années auparavant l'organisation terroriste État islamique (ISIS) avait, par exemple, mené une vaste campagne sur Twitter pour diffuser une propagande visant à encourager la radicalisation et à recruter des soldats étrangers pour sa guerre en Irak et en Syrie (Klausen, 2015).

Les activités d'influence lancées par des acteurs gouvernementaux ou non étatiques avaient cours bien avant l'arrivée des médias sociaux, mais, de par leur portée, leur gravité et leurs répercussions, ces activités ont actuellement pris une ampleur sans précédent, qui risque fort de croître au fur et à mesure que les plateformes numériques gagneront en importance sur Internet et joueront un rôle encore plus central dans les sphères sociales, économiques et politiques. De telles manigances représentent un défi évident pour la cybersécurité. Il n'en reste pas moins que les démocraties, qui dépendent du partage libre et ouvert de l'information, sont particulièrement susceptibles d'être contaminées par le poison des activités d'influence qui propagent de fausses nouvelles et de la propagande et sont source de

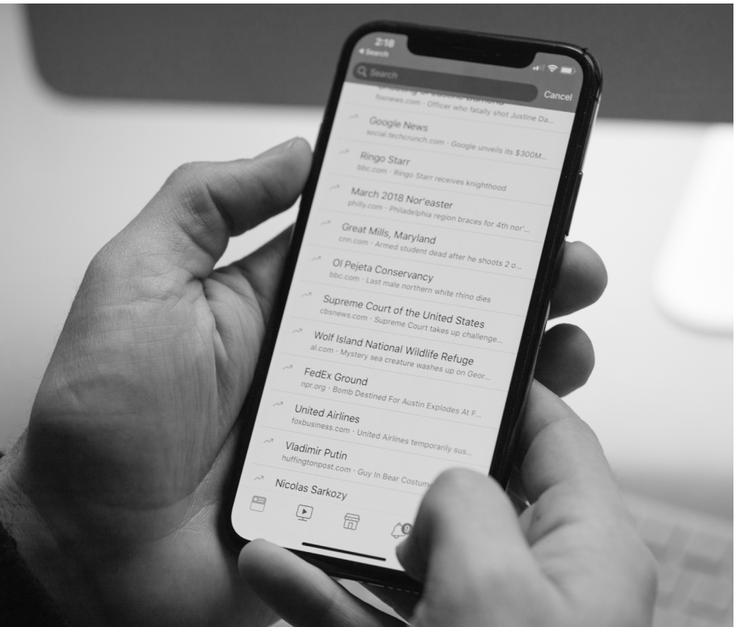
Eric Jardine est professeur adjoint de science politique au Virginia Tech et agrégé au Centre pour l'innovation dans la gouvernance internationale (CIGI). Il axe ses recherches sur les utilisations et les violations du Web invisible; il évalue les tendances de la cybersécurité, la façon dont les gens s'adaptent aux perceptions changeantes du risque lorsqu'ils utilisent de nouvelles technologies de sécurité ainsi que l'environnement politique inhérent qui entoure tant les technologies d'anonymisation que le cryptage. Il a co-rédigé l'ouvrage intitulé *Look Who's Watching: Surveillance, Treachery and Trust Online* (CIGI Press, 2017).

désinformation. La gouvernance démocratique repose entièrement sur le principe selon lequel les citoyens sont bien informés et possèdent non seulement le bon sens nécessaire pour juger de la pertinence des faits et des histoires partagés publiquement, mais également une confiance solide dans l'information communiquée par les institutions. Or, ce système est menacé par des activités d'influence savamment orchestrées qui ne pourront qu'empirer au fur et à mesure que de nouvelles technologies propices aux « fausses nouvelles » entreront en jeu.

La portée du problème

Selon l'un des comptes rendus erronés diffusés en 2016, la candidate du parti démocrate, Hillary Clinton, et son chef de cabinet, John Podesta, étaient à la tête d'un réseau d'exploitation sexuelle d'enfants mené à partir du sous-sol d'une pizzeria de Washington (DC). Ce qui avait débuté sous la forme d'une rumeur malicieuse sur Internet a rapidement pris la forme d'un véritable raz-de-marée sur les médias sociaux. Le mot-clic #pizzagate est devenu viral et des milliers de comptes ont affiché des « preuves » pour et contre cette histoire. Or, bon nombre de ces messages provenaient de l'étranger, majoritairement, et de façon disproportionnée, de la République tchèque, de Chypre et du Vietnam. Peu après l'élection, cette légende en ligne a eu une répercussion sinistre dans le monde physique, lorsque l'un des adeptes de l'histoire, Edgar Welch, s'est rendu en voiture jusqu'à Washington avec un fusil d'assaut. Il est entré dans la pizzeria, a demandé à voir le sous-sol (le bâtiment n'en n'a pas) avant de tirer trois coups. Ainsi, ce qui avait commencé sous la forme d'une désinformation en ligne s'est transformé en véritable drame (Fisher, Cox et Hermann, 2016).

Les activités d'influence existent depuis très longtemps, mais leur portée potentielle est aujourd'hui accrue par l'énorme bassin d'utilisateurs des plateformes et des applications numériques. (Photo : Unsplash.com)



Le scandale du *pizzagate* n'est qu'une seule illustration des problèmes sans cesse croissants que posent les activités d'influence menées par des gouvernements étrangers et des acteurs non étatiques. Bien qu'un écosystème sain s'accompagne de la liberté d'information et d'interprétation des faits, les grandes vagues d'activités d'influence menées à ce jour, plus particulièrement celles axées sur le monde occidental, peuvent être familièrement appelées « fausses nouvelles », car leur contenu est intentionnellement et objectivement faux et peut induire les lecteurs en erreur (Allcott et Gentzkow, 2017, 213). En sus de subvertir les faits, les fausses nouvelles jouent également un autre rôle : elles sont conçues pour trouver écho auprès des lecteurs. Et un tel écho n'est pas produit uniquement par l'information : il peut aussi reposer sur le sentiment qu'il provoque ou le sens qu'a le lecteur de sa véracité, ce qui crée ce qu'on pourrait appeler un élément folklorique (Frank, 2015).

Or, si les fausses nouvelles ne servaient qu'à propager de l'information erronée, alors ceux qui croient de telles histoires seraient soit ignorants, soit ils manqueraient de discernement face aux nouvelles en général, ou alors ils accepteraient de plein gré un contenu incorrect. Mais, si l'on considère les fausses nouvelles comme un genre folklorique, ce que suggère Russell Frank (ibid.), une troisième possibilité apparaît : les fausses nouvelles séduisent, car elles constituent un discours moral ou confirment des sentiments que les gens ont déjà. De ce point de vue, la propagande de l'ISIS sur les médias sociaux concernant la corruption du monde occidental (Klausen, 2015) ou les histoires mensongères sur la santé d'Hillary Clinton durant les élections de 2016 (Milligan, 2016) ont un fondement commun : elles servent à propager une information « alternative » *et* constituent un discours moral auquel les personnes qui ont des points de vue similaires peuvent s'accrocher.

Les activités d'influence qui utilisent des messages aux qualités informationnelles à saveur politique peuvent être lancées par des acteurs étatiques ou non-étatiques ou les deux à la fois. Les efforts visant à influencer les milieux de l'information ont cours depuis longtemps, mais, aujourd'hui, la portée potentielle des activités d'influence est résolument étendue par les nouvelles plateformes numériques qui comptent d'innombrables utilisateurs. À elle seule, la plateforme de Facebook affiche environ 2,25 milliards d'utilisateurs. Twitter en a 336 millions. De même, les applications de messagerie mobile qui permettent aux utilisateurs de partager des fils et des histoires rejoignent une immense proportion des utilisateurs d'Internet : il y a 100 millions d'utilisateurs de Telegram, 1,5 milliard d'utilisateurs de WhatsApp et 1 milliard d'utilisateurs de Viber, sans oublier les plus petites applications de messagerie que l'on trouve en ligne.

L'effet multiplicateur des médias sociaux constitue un tremplin facilement accessible pour les organisations terroristes qui cherchent à radicaliser des personnes ou à recruter des combattants étrangers. Par exemple, l'ISIS a mené en ligne une activité d'influence hautement avancée. Sur Twitter, ce processus s'est répandu géographiquement : des combattants soigneusement sélectionnés en Syrie et en Irak ont affiché sur Twitter des photos qui ont été approuvées et partagées par des tiers et des personnes liées à l'ISIS mais résidant dans le monde occidental (Klausen, 2015). Grâce à cette méthodologie de surveillance simple, l'ISIS a été en mesure de mettre en place une campagne d'influence coordonnée conçue pour donner une image faussement glorieuse de la guerre et de la vie au sein de l'ISIS.

Sur d'autres plateformes numériques, telles que YouTube, l'ISIS a utilisé l'immense bassin d'utilisateurs (1,8 milliard d'utilisateurs) et des heures de visionnement de vidéos (jusqu'à un milliard d'heures chaque jour) pour diffuser des vidéos de propagande visant à glorifier son programme terroriste (Gillespie, 2018). Comme Tarleton Gillespie l'explique, l'ISIS s'est révélée particulièrement habile dans la diffusion de magazines de recrutement sur papier glacé et de vidéos documentant la décapitation de prisonniers politiques et de journalistes (ibid., 55). L'objectif était d'atteindre les personnes susceptibles d'être touchées par les messages de l'ISIS et de les encourager à entreprendre des activités dans leur région ou à aller combattre à l'étranger.

De nouveaux assemblages d'algorithmes socio-techniques favorisent également l'extension de la portée de ces activités d'information. Les bots algorithmiques, qui sont des programmes spécialement conçus pour utiliser la puissance des processus informatiques afin de propager du contenu par l'intermédiaire de faux comptes d'utilisateur, ont favorisé la création de renseignements erronés et la pollution de l'écosystème de l'information en ligne. De tels bots sont particulièrement actifs durant les événements politiques. L'élection américaine de 2016 a connu un revirement en partie à cause de changements quelque peu inattendus dans les préférences de vote au Michigan. Des chercheurs du programme de propagande informatique d'Oxford ont découvert que des nouvelles issues de sources non-professionnelles (fausses nouvelles) ont été partagées plus fréquemment sur les médias sociaux que les nouvelles professionnelles destinées au grand public (Howard et coll., 2017). Plus troublant encore, les nouvelles produites par de grands médias réputés (*The New York Times*, par exemple) ont vu leur pourcentage de contenu partagé chuter à un niveau inférieur record le jour précédent les élections (ibid.). Ces tendances ont été exacerbées par les activités du bot.

La sophistication croissante de l'intelligence artificielle (IA) et des algorithmes de l'apprentissage-machine sont également susceptibles de donner lieu à un nouveau changement dans la qualité des activités d'influence. En règle générale, les gens ont tendance à faire un peu moins confiance à un texte écrit qu'à un message audio et, plus particulièrement, à une vidéo. Une nouvelle peut, certes, indiquer qu'Hillary Clinton est malade, mais cette histoire semblerait plus crédible si M^me^ Clinton l'affirmait ou, du moins, semblait l'affirmer de sa propre bouche. À cet égard, on peut maintenant tirer parti de l'IA pour produire ce qu'on appelle des « contrefaçons profondes », c'est-à-dire des vidéos créées de toutes pièces dans lesquelles une personne annonce une nouvelle erronée (Giles, 2019). Comme il est difficile de déceler les contrefaçons profondes, elles augmenteront d'autant plus les répercussions qualitatives des fausses nouvelles et des activités d'influence étrangères.

En raison de cette portée accrue, de l'augmentation de l'automatisation et du pouvoir des contrefaçons profondes, les activités d'influence des gouvernements étrangers et des acteurs non étatiques ont pris une nouvelle dimension. Des activités gérables à l'ère prénumérique posent aujourd'hui de véritables défis aux régimes démocratiques et libéraux.

Le défi

Fondamentalement, la démocratie repose sur la confiance : la confiance réciproque, la confiance dans les institutions et la confiance dans la crédibilité de l'information. Les activités d'influence, plus particulièrement celles menées par des gouvernements étrangers ou des acteurs non étatiques malicieux peuvent polluer un milieu informationnel, et ainsi saper la confiance des gens et jeter le trouble dans le débat public.

Le discours entourant l'élection présidentielle américaine de 2016 en est un exemple. Durant

Les bots algorithmiques, qui sont des programmes spécialement conçus pour utiliser la puissance des processus informatiques afin de propager du contenu par l'intermédiaire de faux comptes d'utilisateur, ont favorisé la création de renseignements erronés et la pollution de l'écosystème de l'information en ligne.

Les nouvelles produites par de grands médias traditionnels réputés ont vu leur pourcentage de contenu partagé chuter à un niveau inférieur record le jour précédent les élections américaines de 2016. (Photo : Osugi / Shutterstock.com)

la campagne, le débat a été marqué par un degré de rancœur élevé. Depuis les élections, des répondants à un sondage ont indiqué qu'ils avaient l'impression que la civilité et la confiance dans de grandes institutions des États-Unis avaient diminué au fur et à mesure que les camps aux idéologies opposées durcissaient leurs positions. Par exemple, à l'issue d'un sondage, on a découvert que moins de 30 % des personnes faisaient confiance aux institutions médiatiques et, plus largement, non moins de 70 % des répondants avaient le sentiment que la courtoisie était moins présente (Santhanam, 2017).

En outre, des « bulles de filtres » rendent les fausses nouvelles et les autres activités d'influence plus puissantes (Pariser, 2012). Ces bulles de filtres sont le résultat de machinations algorithmiques qui orientent les gens vers des écosystèmes relativement clos d'information en ligne de leur propre cru. Une fois aspirés dans une telle bulle, les gens reçoivent plus d'information qui leur plaît, en fonction des choix qu'ils ont faits auparavant en ligne, qu'il s'agisse de vidéos de chats amusantes sur YouTube ou de messages et de balados à saveur idéologique. Ce qui est troublant est que les filtres des plateformes, dont l'objectif commercial est de donner aux gens ce qu'ils veulent pour les encourager à consommer du contenu, semblent ne pas fonctionner correctement dans l'espace politique, où ils font en sorte que les gens entendent leur propre message, et non pas le point de vue d'autres personnes, dans une chambre de résonance renforcée par des algorithmes. Ainsi, alors que la démocratie exige le libre échange d'idées et d'information, les bulles de filtres tendent à isoler les utilisateurs. En effet, l'information ne peut pas circuler largement et librement dans un milieu filtré.

Bien que les gouvernements soient limités dans leur capacité de restreindre l'exposition, les entreprises auxquelles appartiennent les plateformes ne le sont pas.

Solutions et marche à suivre

Les activités d'influence malicieuses sont un problème croissant exacerbé par les plateformes des médias sociaux, qui permettent d'étendre la portée de la désinformation, de la propagande et des activités de perturbation de l'information ainsi que par les nouvelles technologies algorithmiques susceptibles de nous faire perdre confiance, même en ce que nous voyons de nos propres yeux.

Des changements modestes, mais réels, sont possibles et nécessaires. Globalement, pour lutter contre ce problème, il faut traiter trois aspects : l'exposition, la réceptivité et le contre-discours.

L'exposition est au cœur du problème des fausses nouvelles et des autres formes d'activités d'influence. En effet, une personne peut être psychologiquement mûre pour la radicalisation, mais si elle n'est pas exposée au message de l'ISIS, elle pourrait ne jamais passer à l'acte. De même, l'exposition de l'électorat à de fausses nouvelles durant un cycle électoral peut influer sur le discours politique et même sur les résultats de l'élection. En d'autres termes, en réduisant l'exposition aux activités d'influence, on peut réduire les effets de ces activités.

Dans les démocraties libérales, où la liberté d'expression fait partie intégrante des droits fondamentaux, les gouvernements ne peuvent souvent pas censurer directement l'information partagée en ligne. De plus, l'infrastructure primaire qui sert à diffuser l'information durant une activité d'influence (comme les plateformes des médias sociaux) appartient à des entreprises privées. Cependant, bien que les gouvernements soient limités dans leur capacité de restreindre l'exposition, les entreprises auxquelles appartiennent les plateformes ne le sont pas. Facebook, Twitter et YouTube (exploités par Alphabet) peuvent, en effet, tous contrôler directement le type d'information qui passe sur leur réseau.

Bien que les plateformes aient toujours évité la modération explicite de leur contenu et, dans une certaine mesure, l'évite encore, alléguant qu'elles ne sont pas des éditeurs, les consommateurs ont commencé à exprimer leur désir de voir une certaine modération concernant les contenus plus extrêmes et polarisants, comme celui des suprémacistes blancs et les fausses nouvelles. Les plateformes sont en mesure de le faire (Gillespie, 2018). Elles peuvent, en effet, modérer et ainsi contrôler l'exposition à l'information à l'aide de deux méthodes complémentaires. Premièrement, elles peuvent tirer parti de leur vaste bassin d'utilisateurs en encourageant ces derniers à signaler le contenu potentiellement répréhensible. Elles peuvent ensuite évaluer ce contenu. Si celui-ci est jugé contraire aux conditions de service ou aux directives communautaires de la plateforme, elles peuvent l'éliminer et interdire le compte l'ayant affiché (ibid.). En sus de ces méthodes humaines, bon nombre d'entreprises utilisent des systèmes de détection automatisés pour déceler et éradiquer du contenu. Avec l'arrivée de données supplémentaires, ces approches continueront de s'améliorer. Grâce à ces deux mesures, les plateformes peuvent limiter les pires effets des activités d'influence malicieuses en réduisant l'exposition à du contenu tel que les vidéos de décapitation et de conspiration ainsi que les gazouillis haineux.

Un autre moyen de contrer les activités d'influence est de renforcer l'immunité des personnes en ligne pour qu'elles soient moins réceptives à l'information trompeuse, fausse et polarisante. Les vastes initiatives éducatives qui visent à sensibiliser les utilisateurs aux contenus erronés peuvent être utiles, mais elles sont terriblement dispendieuses. L'inoculation de points clés (personnes) au sein d'un réseau est probablement plus efficace et moins cher (Christakis et Fowler, 2011). La mobilisation ciblée de personnes au centre de réseaux (résultats élevés de place centrale du réseau, en termes d'analyse de réseau social) pourrait favoriser l'immunité de la population et réduire la réceptivité aux faux contenus (Halloran et coll., 2002).

Finalement, les gouvernements et les institutions médiatiques traditionnelles peuvent créer leurs propres exposés des événements pour contrer les activités d'influence. Cependant, l'efficacité de ces contre-discours dépend de la confiance que les utilisateurs accordent à leurs sources; il est donc essentiel de les publier rapidement pour endiguer le flot des activités d'influence perturbatrices visant à saper la confiance des utilisateurs. Leur efficacité dépend aussi probablement de la façon dont les producteurs traditionnels arrivent à s'adapter à l'évolution des médias. L'écosystème de réseautage social actuel est mu par des pièges à clics. L'envoi de titres ennuyeux dans ce type de tourbillon est voué à l'échec.

Cependant, si l'on cible judicieusement les messages des activités d'influence étrangères par un discours inverse, il est possible d'exercer un effet positif sur la perception des utilisateurs d'Internet. Par exemple, la mesure dans laquelle le public est disposé à croire des histoires niant le changement climatique diminue si l'on répond rapidement à ce type de désinformation par des discours contraires qui mettent en lumière les lacunes de la science anti-changement climatique et soulignent le consensus qui existe au sein du milieu scientifique concernant le changement climatique (Cook, Lewandowsky et Ecker, 2017). En bref, bien que réfuter la désinformation soit une bataille continue et non pas une victoire rapide, les gouvernements peuvent, avec l'aide des plateformes, contrer une activité d'influence par une autre. Ce faisant, ils peuvent aider à préserver la confiance ce que suggère la liberté d'information, qui constitue le cœur de la gouvernance démocratique.

Conclusion

Les activités d'influence qui ciblent les régimes démocratiques et libéraux sont profondément troublantes. Elles perturbent les deux piliers d'une gouvernance démocratique efficace : la liberté d'information et la confiance. Ces campagnes peuvent être lancées par des gouvernements étrangers malicieux qui souhaitent semer le chaos, ou par des acteurs non étatiques, tels que l'ISIS, qui cherchent à radicaliser des rebelles du monde occidental. Il est nécessaire, et possible, de contrer ces activités. De telles initiatives nécessitent toutefois la participation non seulement des gouvernements, mais aussi des plateformes. S'ils unissent leurs forces, ces acteurs peuvent préserver la gouvernance démocratique libérale en minimisant l'exposition aux fausses nouvelles et aux autres activités d'influence, en favorisant l'immunité des utilisateurs et en promulguant des contre-discours à la désinformation.

Ouvrages cités

Allcott Hunt et Matthew Gentzkow. 2017. « Social Media and Fake News in the 2016 Election ». *Journal of Economic Perspectives* 31 (2): 211–36. https://web.stanford.edu/~gentzkow/research/fakenews.pdf.

Christakis Nicholas A. et James H. Fowler. 2011. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives — How Your Friends' Friends' Friends Affect Everything You Feel, Think, and Do*. New York, NY: Back Bay Books.

Cook John, Stephan Lewandowsky et Ullrich K. H. Ecker. 2017. « Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence ». *PloS ONE* 12 (5): e0175799. https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0175799&type=printable.

Fisher Marc, John Woodrow Cox et Peter Hermann. 2016. « Pizzagate: From rumor, to hashtag, to gunfire in D.C. ». *The Washington Post*, le 6 décembre.

Frank Russell. 2015. « Caveat Lector: Fake News as Folklore ». *The Journal of American Folklore* 128 (509): 315–32. doi:10.5406/jamerfolk.128.509.0315. www.researchgate.net/publication/281601869_Caveat_Lector_Fake_News_as_Folklore.

Giles Martin. 2019. « Five emerging cyber-threats to worry about in 2019 ». *MIT Technology Review*. www.technologyreview.com/s/612713/five-emerging-cyber-threats-2019/.

Gillespie Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Halloran M. Elizabeth, Ira M. Longini Jr., Azhar Nizam et Yang Yang. 2002. « Containing Bioterrorist Smallpox ». *Science* 298 (5597): 1428–32. http://science.sciencemag.org/content/298/5597/1428.

Howard Philip N., Gillian Bolsover, Bence Kollanyi, Samantha Bradshaw et Lisa-Maria Neudert. 2017. « Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter? » *Data Memo 2017.1*, le 26 mars. Oxford, R.-U.: Project on Computational Propaganda. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf.

Klausen Jytte. 2015. «Tweeting the *Jihad*: Social Media Networks of Western Foreign Fighters in Syria and Iraq ». *Studies in Conflict & Terrorism* 38 (1): 1–22. www.tandfonline.com/doi/abs/10.1080/1057610X.2014.974948.

Milligan Susan. 2016. « Hillary's Health: Conspiracy or Concern? » *U.S.News*, le 15 août. www.usnews.com/news/articles/2016-08-15/hillarys-health-conspiracy-or-concern.

National Intelligence Council. 2017. *Assessing Russian Activities and Intentions in Recent US Elections*. Office of the Director of National Intelligence, Intelligence Community Assessment 2017-01D, le 6 janvier. www.dni.gov/files/documents/ICA_2017_01.pdf.

Pariser Eli. 2012. *The Filter Bubble: What the Internet Is Hiding from You*. Londres, R.-U.: Penguin Books.

Santhanam Laura. 2017. « New poll: 70% of Americans think civility has gotten worse since Trump took office ». PBS News Hour, le 3 juillet. www.pbs.org/newshour/politics/new-poll-70-americans-think-civility-gotten-worse-since-trump-took-office.