

le contre-discours un examen du contenu contestant l'extrémisme en ligne

Jamie Bartlett
Alex Krasodonski-Jones

Octobre 2015

Open Access. Some rights reserved.

As the publisher of this work, Demos wants to encourage the circulation of our work as widely as possible while retaining the copyright. We therefore have an open access policy which enables anyone to access our content online without charge. Anyone can download, save, perform or distribute this work in any format, including translation, without written permission. This is subject to the terms of the Demos licence found at the back of this publication. Its main conditions are:

- Demos and the author(s) are credited
- This summary and the address www.demos.co.uk are displayed
- The text is not altered and is used in full
- The work is not resold
- A copy of the work or link to its use online is sent to Demos.

You are welcome to ask for permission to use this work for purposes other than those covered by the licence. Demos gratefully acknowledges the work of Creative Commons in inspiring our approach to copyright. To find out more go to www.creativecommons.org



Published by Demos October 2015

© Demos. Some rights reserved.

Unit 1, Lloyds Wharf
2-3 Mill Street
London
SE1 2BD

hello@demos.co.uk

www.demos.co.uk

This paper was supported by Facebook. The views expressed are those of the authors and do not necessarily reflect those of Facebook. All errors and omissions are the authors' own.

CONTENU

Introduction	4
Méthode	6
Observations	8
Recommandations	14
Pistes de réflexion	15
Notes	18

INTRODUCTION

Facebook est utilisée par près de 1.5 milliard de personnes autour du monde. Si la majorité d'entre eux l'utilisent bienveillamment, il y en a parmi eux qui utilisent cette plateforme à de mauvaises fins.

De ce fait, Facebook a créé un ensemble de conditions et de règlements, les *Standards de la communauté*, détaillant les contenus publiables et non-publiables sur ce réseau social. Par exemple, Facebook interdit et supprime les propos haineux, définis comme un 'contenu qui cible des personnes en raison de leur race, leur ethnicité, leur origine nationale, leur religion, leur sexe, leur orientation sexuelle, leur infirmité, ou leur état de santé réel ou supposé'. Alors que Facebook proscriit les propos haineux, il arrive que des contenus désagréables ou perturbants qui n'enfreignent tout de même pas les conditions d'utilisation de Facebook soient publiés.

Pour contrer ce genre de contenu extrémiste ou désagréable, Facebook a publiquement déclaré que la pratique du contre-discours, ainsi que les outils fournis pour son développement, doivent jouer un rôle critique. Selon Facebook, cette méthode n'est pas seulement probablement la plus efficace d'adresser l'extrémisme en ligne, mais a également de meilleures chances d'y succéder à long-terme.

L'utilisation du contre-discours est une réaction participative et couramment employée contre l'extrémisme et le contenu haineux. En effet, les publications Facebook à contenu extrémiste sont souvent reçues avec désaccord et dérision, ainsi qu'avec des contre-campagnes. Il est avantageux de combattre l'extrémisme ainsi puisque cette méthode est plus rapide, plus flexible, responsive et capable de le combattre n'importe où et dans n'importe quelle langue – tout en promouvant le principe d'espaces publics ouverts à la libre expression. De plus, les formes prises par ce contre-discours sont toutes aussi variées que l'extrémisme auquel elles s'opposent. Elles ne sont pas toujours aussi efficaces qu'elles ne pourraient l'être, et il faut également considérer la possibilité que le contre-discours pourrait éventuellement être contre-productif.

Motivés par leur forte conviction du pouvoir du contre-discours ainsi que leur intérêt croissant pour une approche plus rigoureuse et scientifique afin de mieux comprendre ce phénomène, Facebook a demandé à Demos d'entreprendre une série de projets de recherche, examinant la mesure dans laquelle différents types de contre-discours sont publiés et partagés sur Facebook.

Ce court rapport intermédiaire décrit et résume les observations du premier stage du projet. Dans ce stage, nous avons examiné comment les discours qui contestent les pages populistes de droite en Europe sont produits et partagés. Les rapports qui

suivent examineront les discours et les contenus qui combattent les idéologies extrémistes islamiques, au Royaume-Uni et autre part.

Il est extrêmement difficile de créer une définition objective d'une *page haineuse*, et nous ne souhaitons pas indiquer que les pages incluses dans cette étude, ou encore leur contenu, soit *haineux*. Plutôt, nous avons choisi d'observer les pages Facebook de groupes populistes de droite, qui sont souvent accusées de donner lieu à et de partager un nombre élevé de propos haineux. Durant cette étude, nous nous référons à ces pages Facebook comme des *pages populistes de droite*. Nous y avons trouvé une grande variété de contenu publié et partagé par ces pages. Afin de choisir les pages représentant le contre-discours, nous avons tenté d'identifier explicitement les pages pour lesquelles une part importante de leur identité comprenait une opposition ou une réponse aux groupes, aux pages et aux contenus qu'ils considèrent haineux.

Nous tenons comme essentiel que le principe de liberté de l'internet doit être maintenu, et qu'il devrait y avoir un espace public où il est possible de s'exprimer librement. Nous tenons d'avantage que le débat, le désaccord et la contestation sont presque toujours préférables à la censure et la suppression des contenus, même lorsqu'il s'agit de contenu extrémiste ou radicalisé quelque soit son origine. Pour autant, nous sommes convaincus que cette idée peut et doit être examinée empiriquement afin de nous aider à mieux comprendre ce phénomène du discours haineux ainsi que la meilleure manière de le combattre. C'est là l'ambition de ce projet de recherche.

MÉTHODE

En utilisant l'*API* publique de Facebook (Application Programming Interface), nous avons recueilli les publications et les données d'interaction de 150 pages Facebook publiques représentant des groupes populistes de droite ou des groupes de contre-discours au Royaume-Uni, en France, en Italie et en Hongrie. Nous avons utilisé 'R' un programme open source qui permet aux chercheurs d'avoir accès aux données publiquement accessible de l'*API* de Facebook.

Nous n'avons pas tenté de recueillir ou d'utiliser des informations personnelles ou individuelles. Nous n'avons pas non plus tenté d'identifier les individus participant à ces conversations. Si le nom ou l'identité d'un individu a été recueilli par erreur, nous l'avons aussitôt supprimé. Nous n'avons pas recueilli de données provenant de pages de groupes ou d'individu ni de pages secrètes ou fermées. Pendant le projet, uniquement les données de pages publiques et visibles à tous ont été utilisées. Afin de particulièrement protéger la vie privée des individus, nous n'avons ni cité ni republié les publications qui auraient pu permettre d'identifier leurs auteurs.

Les pages spécifiques ont été choisies par des chercheurs de Demos, et ne devraient pas être considérées comme représentant un échantillonnage complet des pages pertinentes. Ceci n'est pas une manière idéale d'identifier les pages pertinentes, et nous souhaitons souligner qu'il s'agit d'une étude pilote. Les publications à venir développeront cette méthodologie de manière plus approfondie.

Au cours de deux mois (du 1^{er} Octobre au 1^{er} Décembre 2014), nous avons recueilli 27 886 échantillons publiés par 150 pages au Royaume-Uni, en France, en Italie et en Hongrie. (Par contre, la plupart de l'analyse concerne le Royaume-Uni, la France et l'Italie). Par *publications*, nous nous référons aux mises à jour publiées par les administrateur(s) de ces pages. En plus des publications, nous avons recueilli toutes les interactions liées à ces publications. Les *J'aime*, les commentaires et les partages comptent tous comme une interaction. Les données d'interaction sont utiles pour estimer la portée d'un certain contenu, parce que chaque fois qu'un utilisateur interagit avec un certain contenu, ce dernier apparaîtra sur les *journaux* de leur amis (en dépendent des paramètres de confidentialité choisis par l'utilisateur). En tout, nous avons relevé 8.4 million d'interactions.

Ces données ont été utilisées pour une série d'analyses, dont: le calcul d'un nombre moyen d'interactions utilisant les résultats *API* automatique ; le calcul du format le plus populaire de type de données utilisant les résultats *API* automatiques, le calcul du type et du style des publications les plus populaires réalisé par une analyse humaine manuelle ; le calcul du type de réseau à l'intérieur de ces pages, réalisé par un analyseur automatique de réseaux ; le calcul des différents types de contenu

partagés sur ces pages par rapport au contenu du journal de l'utilisateur réalisé par une analyse automatique.

Il est important de souligner que dans la plupart des cas ces méthodologies sont plutôt expérimentales, car il n'y a pas de méthode fermement établie comme pratique exemplaire pour recueillir et analyser ce type d'information. De plus, il s'agit d'une étude exploratoire. Il faudra donc lire les conclusions avec précaution.

OBSERVATIONS

Données globales sur l'importance et la portée des discours haineux ou contre-discursifs

En total, nous avons identifié 124 pages populistes de droite réparties entre quatre pays, en comparaison à 26 pages contre-discursives (pour les raisons expliquées ci-dessus). De même, il y avait beaucoup plus de publications réalisées par les pages populistes de droite (25 522) que sur les pages contre-discursives (2 364).

Évidemment, ces pages avaient également beaucoup plus d'interactions. Les pages populistes de droite comptaient 7,8 millions d'interactions, comparé à 546 000 interactions pour les pages contre-discursives.

La méthode utilisée pour le recueil de données résulte en la prépondérance de données de pages populistes de droite que de pages contre-discursives. Ceci ne signifie pas nécessairement qu'il en existe plus, mais plus probablement que les pages contre-discursives sont plus difficiles à identifier facilement. Pour présenter des chiffres plus représentatifs, nous avons choisis d'utiliser le nombre moyen de publications ainsi que le nombre moyen d'interaction par publication dans chaque pays et pour chaque type de page. Ceci nous a permis de trouver que généralement, au Royaume-Uni, les pages contre-discursives sont moins nombreuses et moins actives que les pages populistes de droite. Par contre, les publications contre-discursives obtiennent un nombre plus important de partages et d'interactions que celles des pages populistes de droite.

Tableau 1 Nombre moyen d'interactions par publications par pays

Pays	Contre-discours	Populiste de droite
France	126	285
Italie	37	235
Royaume-Uni	402	325

Analyse des réseaux d'adhérents et de leur rôle dans la dissémination du contenu

Afin de mieux comprendre la manière dont l'information et les idées se répandent d'un page à l'autre, nous avons sélectionné les pages provenant du Royaume-Uni et nous avons examiné le degré auquel les contributeurs d'une page avaient également interagi avec une autre page du même type. Ces informations ont été recueillies grâce aux système publique *API* de Facebook ainsi que le logiciel *R* et visualisé dans *Gephi*, un outil d'analyse open source.

Pour les pages populistes de droite, nous avons analysé 92 pages. Un total de 54 495 utilisateurs uniques ont contribué à ces pages, qui entre eux ont également écrit 150 437 commentaires sur ces pages pendant les deux années étudiées. (Ce résultat a été obtenu en recueillant les commentaires et en identifiant le nombre d'utilisateurs uniques ayant contribué à cette série de données.) Nous avons trouvé que 16,2 % d'utilisateurs sont actifs sur plus d'une page. De plus, 1,3 % d'entre eux sont actifs sur plus de trois pages.

Nous avons ensuite analysé 21 pages contre-discursives. Ici, un total de 116 534 utilisateurs uniques ont contribué à ces pages, comptant entre eux 135 842 interactions. Malgré le fait qu'elles ont plus d'utilisateurs, les réseaux de ces pages sont moins denses : 12,7% d'utilisateurs sont actifs sur plus d'une page et 1.15% d'entre eux sont actifs sur plus de trois.

Ceci indiquerait qu'au Royaume-Uni, les pages populistes de droite ont un réseau un peu plus concentré, avec moins d'utilisateurs plus actifs. En contraste, les pages contre-discursives ont plus d'utilisateurs actifs dont chacun contribue moins de contenu. Cependant, cette étude a besoin d'être répliquée à une échelle plus large afin de produire des résultats significatifs.

Quel type de publication atteint une portée importante de manière efficace

Afin de déterminer quel type de publication parvient à atteindre une portée importante, nous avons examiné a) le format des publications et b) le contenu et le style de ces publications. Les données *API* de Facebook permettent aux chercheurs de déterminer quel format les publications vont prendre, divisées entre *lien*, *photo*, *statut* ou *vidéo*. Le tableau 2 ci-dessous indique quel type de format est utilisé le plus fréquemment par les pages en question.

Tableau 2 Type de contenu par pays (arrondi au pourcentage le plus proche)

	Pays	Lien	Photo	Statut	Vidéo
Contre-discours	France	64	15	14	6
	Hongrie	68	5	26	2
	Italie	29	45	17	9
	Royaume-Uni	40	34	18	8
Populiste de droite	France	26	43	17	14
	Hongrie	10	83	2	5
	Italie	59	27	9	5
	Royaume-Uni	57	23	10	10

On lit sur ce tableau une grande disparité entre les différents pays et les types de pages sur le choix du format des publications à partager. Par exemple, en France, ce

sont les liens qui sont les plus populaires parmi les pages de contre-discours, alors qu'en Italie il s'agissait des photos.

Le niveau auquel un certain type de format est meilleur qu'un autre dépend probablement de plusieurs facteurs. Cependant, en calculant de nombre moyen d'interaction pour chaque type de format, il est possible de tirer quelques conclusions : en général, l'utilisation de photos est la manière la plus effective de générer une interaction avec les utilisateurs (voir tableau 3 ci-dessous).

Tableau 3 Type de contenu utilisé le plus fréquemment par Pays

	France	Hongrie	Italie	Royaume-Uni
Populiste de droit	Photo	Vidéo	Photo	Photo
Contre-discours	Photo	Photo	Photo	Photo

Analyse du contenu des publications

Afin d'obtenir une compréhension plus nuancée du type et du style des publications les plus populaires, nous avons analysé 1 000 publications de pages populistes de droite sélectionnées au hasard (partagées entre la France, l'Italie et le Royaume-Uni) ainsi que 1000 publications de pages de contre discours également sélectionnées au hasard (de même partagées entre la France, l'Italie et le Royaume-Uni). Nous les avons réparties entre diverses catégories décrivant le *contenu* de la publication ainsi que son *style*. Les catégories ont été sélectionnées par les chercheurs, voir les descriptions accompagnées d'exemples ci-dessous. La popularité de chaque publication est estimée en utilisant le nombre d'interactions reçues.

Nous avons trouvé que le contenu des publications des pages populistes de droite était majoritairement commentatif, selon notre classification (en moyenne 2 542 interactions).¹ Ce type de publication reçoit environ double le nombre d'interactions par rapport aux autres types de contenu.

Le contenu le plus populaire parmi les pages de contre discours était interrogatif (en moyenne 10 934 interactions, légèrement biaisé par un petit nombre de publications extrêmement populaires.) Le second type de contenu le plus populaire était de type commentatif. En France, contrairement aux autres pays, le contenu le plus populaire des pages contre-discursives était agressif.²

La plupart de publications des pages populistes de droite adoptent un style glorificateur³, par exemple une publication qui commémore les soldats morts au combat ou qui adopte un ton de fierté patriotique (6 607 interactions en moyenne), suivi par un style courroucé⁴ (4 093 interaction).

Les publications les plus populaires parmi les pages contre-discursives dans chacun des pays sont drôles or satiriques⁵ (2 717 interactions).

Afin de mieux analyser les publications les plus populaires, nous avons examiné 100 publications de style satirique publiées par des pages contre-discursives, c'est à dire le style le plus populaire parmi ce type de publications. Cette analyse a été conduite seulement en Anglais.

En terme d'interactions, les sujets comme l'immigration, l'ethnicité et la religion ont chacun reçu, en moyenne, plus de 4 000 interactions et constituent le type de contenu le plus populaire. Ces publications parodient souvent les discours extrémistes habituellement utilisé par les pages haineuses à ces sujets. La démarche d'un régiment militaire pour demander à Facebook de supprimer une image d'eux publié par *Britain First* (Le Royaume-Uni d'abord, la plus importante page populiste de droite au Royaume-Uni) était l'incident isolé le plus populaire. Une capture d'écran de cette requête a été partagée plus de 9 000 fois.

Analyse du type et de la nature des commentaires

Contrairement aux publications, qui sont publiées par les administrateurs de page, n'importe qui peut publier un commentaire en réponse à une publication. Nous avons analysé 1 000 commentaire sélectionnés au hasard écrits en réponse à des mises à jour (également sélectionnées au hasard) de pages populistes de droite (réparties entre la France, l'Italie et le Royaume-Uni) et 1000 commentaires sélectionnés de la même manière parmi les pages contre-discursives. Les catégories utilisées ont été sélectionnées par nos chercheurs. La popularité d'un commentaire est jugée par le nombre de *j'aime* reçus (vu que c'est le seul type d'interaction permis avec un commentaire).

Sur les pages populistes de droite, 9% des commentaires sont des commentaires contre-discursifs, c'est à dire des commentaires qui sont en désaccord avec la publication ou qui tentent d'y présenter en réponse un message alternatif plus positif. (En France, ce type de commentaire comptait 17% de tous les commentaires, 7% en Italie et 5% au Royaume-Uni). Si on peut généraliser cet échantillon, ceci voudrait dire qu'il aurait 25 000 commentaire contre-discursifs écrits sur des pages populistes de droite chaque mois au Royaume-Uni.

Sur les pages contre-discursives, le contre-discours constructif semble être le type de commentaire le plus populaire (en moyenne 5.9 *j'aime* par commentaire)⁶, suivi de près par le commentaire discursif et constructif (en moyenne 5.3 *j'aime* par commentaire)⁷. Par contraste, les commentaires de contre-discours non-constructifs reçoivent en moyenne 3.3 *j'aime* par commentaire⁸, et les vérifications de faits 3.8

J'aime par commentaire⁹. Par contre, malgré la popularité de ce contenu, le contre-discours constructif ne tient compte que de 6% de tout le contenu analysé, alors que le contre-discours non-constructif en compte 20%.

Afin de mieux comprendre les types de commentaires les plus populaires, nous avons analysés 100 commentaires discursifs et constructifs écrits en Anglais sur des pages de contre-discours (une fois de plus, les catégories ont été sélectionnées par nos chercheurs). Les commentaires adressant une question de politique générale ont le nombre de *j'aime* le plus élevé (13.2), suivi par les commentaires qui visent à interroger un parti ou un détail de programme politique (8.7). Ceci suggère que les commentaires interrogeant des questions politiques, le détail des programmes politiques ou informatifs se répandent plus effectivement. Par contre, ils ne comptent que pour une toute petite proportion du nombre total des commentaires (6% et 3% respectivement).

Analyse de l'origine du contenu partagé et débattu

Nous avons examiné la dissémination de ce contenu entre les journaux des utilisateurs, ainsi que de savoir si la provenance du commentaire affecte la manière dont les utilisateurs y réagissent. Nous avons analysé les interactions de 4 388 publications contre-discursives ainsi que 75 132 publications de pages populistes de droite, et calculé la proportion des publications partagées par la page d'origine par rapport à celles partagées par un autre utilisateur. Pour ceci nous avons utilisé une banque de données plus importante que celle utilisée pour les analyses décrites ci-dessus (les pages utilisées sont les mêmes mais la période de collection de données est plus étendue). La collection de cette deuxième banque de données a été complétée avec l'aide du groupe de Science des Données (*Data Science*) de Facebook et est composée d'informations publiques, collectives, non-expérimentales et historiques.

Tableau 4 Lieu des interactions en réponse aux publications

	Publications populistes de droite			Publications contre-discursives		
	% publiées sur la page d'origine	% publiées par un autre utilisateur	N=	% publiées sur la page d'origine	% publiées par un autre utilisateur	N=
<i>J'aime</i>	94	6	4,138,425	82	19	56,884
Partages	97	3	474,558	99	1	13,430
Commentaires	73	27	2,173,678	60	41	16,232

Ces résultats, exprimés en pourcentage du nombre total de *j'aime*, indiquent que les pages contre-discursives reçoivent plus de *j'aime* et de commentaires lorsqu'une publication est repartagée, c'est à dire de la part des utilisateurs qui ont trouvé le contenu sur le journal d'un utilisateur tiers plutôt que sur la page d'origine (même s'ils *aiment* cette page). Ceci indique que le contenu des pages contre-discursives peut probablement se répandre plus effectivement, et qu'il pourrait éventuellement avoir une portée beaucoup plus importante si plus d'utilisateurs partageaient le contenu avec leurs amis. De plus, nous sommes certains qu'il doit y avoir de nombreux commentaires et discussions sur les partages de ces publications, c'est à dire sur les journaux d'utilisateurs et entre leurs amis. C'est là que nous estimons que la plupart du contre-discours a lieu, mais nous ne pouvons pas accéder à ces discussions pour des motifs reliés au respect de la vie privée.

Utilisant la même méthode, nous avons également examiné comment les différents types de contenu atteignent les utilisateurs qui n'avaient pas *aimé* les pages d'origine. Nous avons trouvé ce résultat en calculant le pourcentage d'utilisateur ayant *aimé* ou commenté sur une publication sans *aimer* la page d'origine.

Tableau 5 La portée d'une publication en dehors des utilisateurs qui *aiment* la page d'origine

	% de <i>j'aime</i> par les utilisateurs n' <i>aimant</i> pas la page d'origine		% de commentaires d'utilisateurs n' <i>aimant</i> pas la page d'origine	
	<i>Populiste de droite</i>	<i>Contre-discours</i>	<i>Populiste de droite</i>	<i>Contre-discours</i>
Liens	50	18	66	32
Photos	52	5	65	25
Vidéos	68	26	75	48
Statut	21	7	41	23

Ceci montre que les pages populistes de droite partagent bien plus efficacement leur contenu dont la portée s'étend bien plus loin que leur réseau d'abonnés. Pour les pages contre-discursives (et les pages populistes de droite) les vidéos sont le type de contenu qui se dissémine le plus efficacement.

RECOMMANDATIONS

En vue des observations ci-dessus, les pages contre-discursives sont moins actives que les pages populistes de droite. Au Royaume-Uni, elles ont un bon nombre moyen d'interactions, ce qui est moins le cas en France ou en Italie. Si ces pages souhaitent avoir une portée plus importante, elles devraient produire plus de contenu. Au Royaume-Uni, les pages contre-discursives ont plus de contributeurs qui demeurent pour autant moins actifs que leurs homologues sur les pages populistes de droite.

Pour ceux qui souhaitent encourager la dissémination du contre-discours, ce projet suggère qu'un recentrage serait profitable à son progrès, plus spécifiquement :

- Les administrateurs devraient publier plus de photos et de vidéos proportionnellement au nombre total de publications.
- Les commentateurs devraient présenter un contre-discours plus constructif autour d'aspect particuliers des programmes politiques.
- Les contributeurs aux pages contre-discursives devraient encourager leurs amis à partager le contenu plus activement.
- Globalement, notre recherche suggère que si les administrateurs et les utilisateurs des pages contre-discursives étaient plus actifs, et changeait légèrement le contenu de leurs publications, ceci pourrait dramatiquement améliorer leur portée.

PISTES DE RÉFLEXION

Le contre-discours est un phénomène bien plus complexe que l'on ne l'avait anticipé. Il est important de comprendre que le contre-discours est bien plus complexe que le simple fait d'être en désaccord ou de contester le contenu d'une publication sur une page publique. Parfois explicite, il prend la forme d'une critique des opinions exprimées sur un journal, ou bien d'utilisateurs qui sont activement à la recherche de contenus haineux afin de relever le défi de front. D'autres réponses sont plus implicites : certains utilisateurs choisissent de signaler le contenu haineux qu'ils rencontrent à Facebook, bloquent ou mettent en sourdine (Facebook *mute*) les utilisateurs qui publient ce genre de contenu, ou encore ils peuvent choisir d'exprimer leur différend dans un message privé. Ailleurs, le contre-discours peut prendre la forme d'un groupe ou d'une page humoristiques en opposition à un individu ou une page.

Parmi ces pages de contre-discours, nous avons trouvé nombre de discussions constrictive et de débat ; mais l'occurrence d'un contre-discours moins constructif est fréquente, tel le recours à des communiqués injurieux ou menaçants. Les catégories que nous avons définies incluent : le contre-discours constructif, la contre-haine non-constructive, l'opposition satirique, la discussion constructive, la vérification des faits et l'opposition réfléchie. Certaines de ces approches méritent d'être privilégiées par rapport aux autres.

Puisqu'il y a tellement de types différents de contre-discours, il est difficile d'estimer ce qui constitue un résultat fructueux. Il n'y a pas de consensus actuel permettant le choix d'indicateurs clés. Nous avons donc utilisé trois différents types d'indicateurs que nous estimons calculable, avec des degrés variés de difficulté).

Mesures quantitatives

L'analyse quantitative de contenu de réseau sociaux est récemment devenu une grande entreprise. Des centaines de compagnies ont construit des milliers d'outils tentant de quantifier l'influence des médias sociaux. Si la plupart de cette analyse est inadéquate à l'analyse scientifique; plusieurs des notions centrales peuvent être mises en pratique pour ce projet.

- **L'engagement est une mesure d'interaction entre utilisateurs.** Par exemple, on peut choisir le taux d'utilisateurs qui ont vu la page par rapport à ceux qui se sont inscrits, ou le nombre d'utilisateurs qui ont vu un morceau de contenu sans l'*aimer* ou le partager. Si cet outil est un peu rude, il peut servir de substitut utile pour estimer la portée possible d'une publication. Cette mesure est facile à utiliser effectivement.

- **Volume et distribution.** Combien de publications sont produites pour chaque thématique, et combien d'utilisateurs uniques en discutent? Il y a-t-il des pics significatifs où la densité du discours est plus importante? Ceci peut être réalisé avec accès à des données relatives au trafic plutôt que des données tirées de pages spécifiques.
- **Mesures de la portée d'une conversation sur un réseau social.** Quelle est la portée réelle de ces conversations? Le discours haineux se limite-t-il à des communautés isolées (que ce soit des communautés qui s'isolent elles-mêmes ou qui se retrouvent isolées par les algorithmes de personnalisation de Facebook)? Avec quelle fréquence ce contenu déborde-t-il dans les journaux d'utilisateurs en dehors de ces communautés? Le calcul de cette portée peut être un outil précieux en combinaison, par exemple, avec une mesure d'engagement. Au cours de ce projet nous avons découvert un nombre d'outils utiles.

Mesurer le contenu

Toute mesure de contenu se base sur un jugement du type et de la nature de ce contenu plutôt que de son volume. Ceci requière souvent une analyse manuelle ou un recours à un outil de traitement automatique du langage naturel (les deux sont possibles – à des degrés variables d'exactitude). La facilité d'échelonner et d'étendre rapidement cette approche à des volumes de contenus beaucoup plus importants est un de ses points les plus forts.

- L'analyse du sentiment est une technique qui accompagnée d'un cadre analytique, nous a permis de déterminer si un contenu peut être classifié comme discours haineux, contre-discours ou encore ni l'un ni l'autre. Il est peut-être possible de juger le degré d'extrémisme d'un contenu en utilisant le langage employé. Nous estimons que ce procédé est difficile mais pas impossible.
- Plus spécifiquement, on devrait pouvoir séparer le contenu haineux et le contre-discours en catégories plus précises en utilisant un outil de traitement automatique du langage naturel : l'incitation à une action hors-ligne; un débat réfléchi, une insulte, etc. Nous estimons que ce procédé est difficile mais pas impossible.
- Il reste à comprendre la mesure dans laquelle une discussion sur Facebook évolue ou se résout : la publication d'un contenu contre-discursif a-t-elle un impact sur le reste de la discussion? Nous estimons qu'il est très difficile de répondre à cette question en utilisant des outils automatiques, mais pas impossible à résoudre manuellement par une analyse humaine.

Mesures réelles

Il s'agit ici d'estimer si la présence du contre-discours en ligne a un effet à long terme sur les attitudes et les comportements hors-ligne. Il est presque impossible de calculer ceci avec exactitude – tout comme il est extrêmement difficile de déterminer

si le contenu haineux en ligne entraîne une modification d'attitudes et de comportements hors-ligne. Le choix d'une mesure pertinente et justifiable à cet égard requiert plus de recherche.

Cette étude exploratoire a également permis d'identifier les domaines où un effort relativement modeste en recherche permettrait une meilleure compréhension du sujet.

Tout d'abord, il faut créer de meilleurs outils pour mesurer le volume total de contenu populiste de droite ou contre-discursif. Il est extrêmement difficile de produire des chiffres globaux et fiables des différents types de contre-discours utilisant une méthode scientifique rigoureuse et objective permettant de définir les pages que nous avons choisi. Avec nos méthodes ad hoc d'identifications, les risques d'obtenir des données erronées sont très élevés. Une approche automatique relevant les pages les plus actives et avec le plus de *j'aime* par pays et par catégorie permettrait d'établir une mesure bien plus robuste et fiable.

Deuxièmement, il y a des nouveaux types de contenu qui pourraient être étudiés dans plus de détail, par exemple : la portée totale des différents types de contenus (plutôt que simplement leurs interactions); l'examen d'une conversation entière en réponse à une publications ainsi que son évolution; s'il y a des groupes ou des utilisateurs particuliers qui ont plus de succès quant à la résolution de disputes ou l'opposition au discours haineux; et l'examen des données des publications partagées sur un journal d'utilisateur (par rapport à la même publication publiée sur la page elle-même).

Troisièmement, il serait utile de comprendre comment Facebook, en tant que réseau d'utilisateur partageant des contenus populistes de droite ou contre-discursifs, peut et doit informer notre méthode. Ceci inclurait l'examen de la manière dont ces contenus se disséminent. En traçant un seul morceau de contenu, par exemple, nous pourrions utiliser une analyse de réseau pour voir: qui le partage, où ils l'obtiennent et où ils l'envoient. Ceci permettrait d'obtenir une esquisse de réponse aux questions fondamentales par rapport à la manière dont ces contenus se transmettent: quels chemins prennent-ils? Différent contenus sont-ils diffusés de la même manière? Ces dynamiques sont elles pareilles ou différentes dans différents pays?

Dernièrement, plus de recherche permettrait de déterminer ce qui pourrait constituer une mesure robuste et pertinente du succès hors-ligne. Une série d'études de cas approfondies des campagnes en ligne ayant eu un impact hors-ligne considérable permettrait de produire un aperçu concret et détaillé de cette mesure.

NOTES

¹ Par commentaire, nous nous référons à tout type de contenu créé ou partagé par la page qui s'adresse à une situation sans nécessairement faire référence à une source extérieure.

² Par agressif nous visions les contenus particulièrement offensifs souvent dirigés vers un groupe ou un individu particulier. Ils peuvent être dirigés à un group ethnique, un politicien ou une organisation.

³ Les contenus glorificateurs sont ceux qui célèbrent la page et ses valeurs.

⁴ Ce contenu peut faire recourt à un registre grossier et faire appel à des mesures extrêmes par rapport au contenu.

⁵ Il s'agit ici de tout le contenu humoristique ou satirique.

⁶ Tout discours tentant d'avoir un débat sérieux sur des sujets spécifiques à propos de discours haineux (racisme, xénophobie et immigration).

⁷ Tout discours tentant d'avoir un débat sérieux sur des sujets généraux de politique, d'affaires actuelles ou d'intérêt général.

⁸ Ce discours conteste les discours haineux, mais de manière non-constructive, par exemple une attaque individuelle ou offensive.

⁹ Cette catégorie vise les commentaires vérifiant ou interrogeant les autres utilisateurs au sujet de fait déclarés ou d'affirmations non-justifiées.

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

- a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.
- b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.
- c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.
- d 'Original Author' means the individual or entity who created the Work.
- e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.
- f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

- a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;
- b to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

- a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicense the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.
- b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital filesharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

C If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

A By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

- i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;
 - ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.
- B except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licensed on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

A This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

B Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

A Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

B If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

C No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

D This Licence constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

Le contre-discours, c'est-à-dire le discours qui conteste, critique ou présente une opinion adverse, sera sans doute un outil important pour contrer l'extrémisme ou les contenus offensifs en ligne. C'est une méthode rapide, flexible et responsive, capable de faire face à l'extrémisme nonobstant sa provenance géographique et dans n'importe quelle langue. C'est également une approche qui maintient le principe de plateformes publiques et ouvertes au débat. Par contre, ce n'est peut-être pas toujours une stratégie aussi efficace qu'elle ne pourrait l'être et certains types de contre-discours pourraient même finir par être contre-productifs.

Ce court rapport intérimaire présente les résultats sommaires d'un nouveau projet de recherche examinant comment les discours dissidents contre les Pages Facebook populistes de droite en Europe sont produits et partagés. Ce rapport détaille les observations tirées de l'examen de l'activité de 150 pages Facebook sur une période de deux mois, et comment ce contenu y est produit et partagé ; ainsi que la manière dont les utilisateurs interagissent avec. Ce projet a également tenté de développer de nouvelles méthodologies et approches pour mesurer la portée du contenu en ligne, son impact et son effectivité.

Il s'agit du premier rapport dans une série de projet examinant le contre-discours et le contenu contestant l'extrémisme en ligne. Les rapports à suivre examineront l'idéologie extrémiste islamiste au Royaume-Uni et autre part.

Jamie Bartlett est directeur du centre pour l'analyse des médias sociaux à Demos. Alex Krasodonski-Jones est chercheur au centre pour l'analyse des médias sociaux à Demos.

Traduit de l'anglais par Gabrielle Cahen

ISBN 978-1-909037-93-9