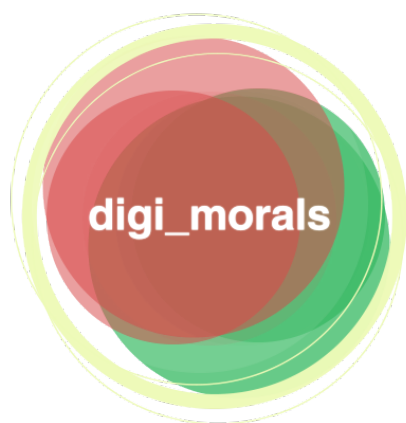




DESACUERDOS EN CONTEXTOS DIGITALES

Variedades, causas y posibles
intervenciones



Antonio Gaitán Torres (coordinador, uc3m)

María Luengo Cruz (coordinadora, uc3m)

Hugo Viciano Asensio (Universidad de Sevilla)

Francisco Seoane Pérez (uc3m)

Teresa Gil López (uc3m)

Javier Lorenzo Rodríguez (uc3m)

Daniel Barbarrusa (Universidad de Sevilla)

Gonzalo Velasco Arias (uc3m)

Fernando Broncano Rodríguez (uc3m)

Gabriela Müggemburg y Rodríguez Vigil (uc3m)

Revisores externos

Manuel Almagro Holgado (Departamento de Filosofía, Universitat de Valencia)

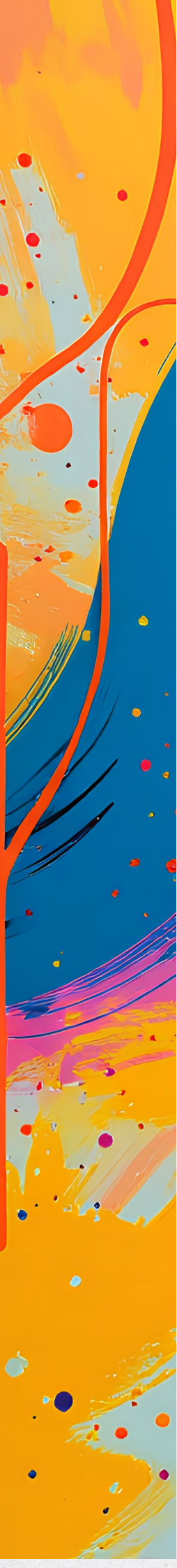
Miriam Juan-Torres González (Othering & Belonging Institute, Universidad de Berkeley)

Luis M. Miller Moya (Instituto de Políticas y Bienes Públicos, IPP-CSIC)

Este documento se ha elaborado durante el desarrollo del proyecto de investigación “Los desacuerdos morales en la esfera digital – dinámicas interactivas, micro-mecanismos y marcadores culturales”.

Este proyecto ha sido financiado por las Ayudas FBBVA a Proyectos de Investigación Científica, convocatoria 2021.

Fundación
BBVA

- 
- Los desacuerdos emergen en nuestras interacciones cotidianas como procesos comunicativos complejos, insertados en dinámicas de intercambio o confrontación de opiniones, con apelaciones a evidencia y marcos evaluativos.
 - Desde una perspectiva epistémica, comunicativa y política resultan especialmente interesantes los desacuerdos entre pares y los desacuerdos profundos.
 - Para entender la incidencia de estos dos tipos de desacuerdo en contextos digitales conviene atender primero a lo que sabemos sobre la diversidad de opiniones online y offline.
 - La evidencia sugiere que los efectos positivos de la diversidad de opiniones no son intrínsecos. Si bien algunas ideologías asumen que la diversidad es inherentemente beneficiosa, hallazgos recientes indican que su impacto positivo depende de factores contextuales clave.
 - Entre estos factores figuran las normas de interacción, la arquitectura de incentivos en los contextos de interacción y las dinámicas preexistentes entre los distintos grupos.
 - Estos mismos principios se aplican a los entornos digitales, aunque requieren una adaptación específica debido a las características únicas de los contextos online. Los nuevos entornos digitales modulan estos factores y añaden otros nuevos.
 - En el entramado de la comunicación digital, los desacuerdos online se manifiestan a través de una serie de rasgos distintivos: 1) las dinámicas de autoselección sustentadas en la arquitectura de las redes sociales; 2) la señalización de afinidad e hiper moralización desde agrupamientos ideológicos e identitarios; y 3) el recurso a marcadores culturales y retóricas de la conspiración para exacerbar la polarización.

Para minimizar los efectos negativos de los desacuerdos en redes sociales y plataformas digitales proponemos cuatro acciones clave:

- Rediseñar las plataformas para reducir la autoselección.
- Aprovechar el potencial de los LLM (“Large Language Model”) para la moderación y explorar los beneficios de chats en contextos reales de desacuerdo (facilitando evidencia en debates concretos, precisando posibles posiciones de consenso, etc.).
- Atender a los marcos narrativos que acentúan el desacuerdo (mediante la exploración de marcadores culturales) y a retóricas de la conspiración que refuerzan la naturaleza grupal y excluyente en algunas interacciones.
- Estimular la creación de formas alternativas de activismo en redes sociales que eviten la confrontación directa en entornos polarizados y fomenten la expresión de la disidencia desde comunidades epistémicas seguras.



Desacuerdos – entre la conexión y la diferencia

Los desacuerdos nos rodean. Desde conversaciones cotidianas hasta debates públicos, las diferencias de opinión moldean nuestra percepción del mundo. Ya sea en el ámbito familiar, entre amigos, en el discurso político o en los medios de comunicación, los desacuerdos son una constante en nuestra interacción cotidiana.

Los desacuerdos, aunque a menudo incómodos, pueden ser una fuente de aprendizaje. Cuando nos enfrentamos a opiniones divergentes, se nos brinda la oportunidad de cuestionar nuestras propias creencias y ampliar nuestra comprensión de la realidad. En un sentido más amplio, las instituciones centrales de nuestra sociedad pueden entenderse como mecanismos para gestionar los desacuerdos y aprovechar su potencial constructivo. Los medios de comunicación y la educación, por ejemplo, son espacios donde se gestionan desacuerdos, buscando extraer valor epistémico de ellos al escuchar y aprender de aquellos con mayor conocimiento o experiencia. Asimismo, los partidos políticos y los diversos mecanismos de representación existentes deberían canalizar nuestros desacuerdos políticos, permitiendo que las diferentes voces sean escuchadas y consideradas en la toma de decisiones.

Sin embargo, los desacuerdos también pueden ser un factor de división cuando intensifican las diferencias y exacerban la distancia entre perspectivas políticas y morales. En el terreno personal, descubrir que un amigo íntimo tiene opiniones opuestas sobre temas como el aborto o la eutanasia puede generar distanciamiento. En el ámbito político, observar que los líderes de los partidos ejemplifican desacuerdos excesivamente intensos sobre cuestiones sociales fundamentales puede tener un impacto negativo en la ciudadanía. La expresión de desacuerdo en la élite incide en la polarización del debate público, pues erosiona la confianza en la neutralidad de las instituciones hacia determinadas posiciones morales.

En **digi_morals** exploramos las dos caras de los desacuerdos, prestando especial atención a los nuevos y complejos contextos digitales.

Nuestro objetivo es comprender a fondo este fenómeno y utilizar las herramientas analíticas y conceptuales que nos ofrecen la Filosofía, la Comunicación y la Ciencia Política para ofrecer un mapa de los diferentes problemas y niveles de análisis.

Gaitán, A. Luengo, M., Lorenzo, J. 2023. ¿Por qué debemos aprender a gestionar los desacuerdos morales en los medios digitales?, *The Conversation*, 13 de noviembre de 2022.

Una estructura comunicativa compleja

La Filosofía se ha preguntado por los desacuerdos desde sus inicios. Los diálogos platónicos comenzaban por un desacuerdo en torno a algún asunto central - la justicia, el conocimiento, la naturaleza del buen gobierno o la belleza. La defensa clásica de la libertad de expresión en John Stuart Mill también apuntaba al potencial que tiene la expresión diversa de opiniones, frente a la censura y la supresión, para exponer los errores sobre un determinado tema. Los desacuerdos han estado en el centro de la filosofía desde sus inicios hasta nuestros días.

Desde hace unas décadas, la epistemología y la filosofía del lenguaje han estudiado los desacuerdos de manera más precisa con el objetivo de entender el potencial que tienen como estructuras comunicativas.

Los desacuerdos son estructuras comunicativas complejas en las que se suelen intercambiar argumentos, razones o evidencia. Ese intercambio de razones, una práctica reglada y sujeta a convenciones culturales, asume que podemos ser sensibles a las buenas razones, cambiando nuestra perspectiva o matizando nuestra convicción.

Conviene distinguir distintas formas de estar en desacuerdo o de intercambiar razones. En lo que sigue destacamos dos tipos de desacuerdo y nos preguntamos por los efectos que tienen los contextos digitales en el potencial epistémico y comunicativo de estos desacuerdos.

Desacuerdos entre pares

Dos personas están en desacuerdo acerca del cambio climático si una afirma que el cambio climático existe y otra niega esa afirmación. Estamos rodeados de desacuerdos de este tipo. En ocasiones resulta fácil decidir qué debemos creer en esas situaciones. Si sabemos que la otra persona no entiende los rudimentos más básicos del clima, esos que enseñan en el instituto, parece sensato seguir creyendo que el cambio climático existe.

La mayoría de desacuerdos, sin embargo, son más complicados. En muchas ocasiones nos encontramos con personas que están igual de informadas que nosotros y que razonan de forma coherente (¡exactamente igual que nosotros!). En esas situaciones resulta más complicado saber qué debemos creer. Si alguien tiene la misma información que yo sobre el cambio climático y es igual de razonable a la hora de evaluar esta evidencia, ¿qué peso debo darle a su opinión contraria?

Para algunos filósofos/as, lo adecuado en estos casos sería suspender la creencia o al menos disminuir la convicción con la que nos comprometemos con la creencia. Los desacuerdos entre pares nos dan la oportunidad de ser humildes sobre nuestras creencias.

Christensen, David. (2007). Epistemology of Disagreement: The Good News. *Philosophical Review*, 116(2), 187–217.

Para otros filósofos/as, en estos casos debemos seguir afirmando nuestra posición. Quizás en este caso la otra persona no ha considerado de forma adecuada la evidencia compartida, o puede que justo hoy ande distraída o se haya equivocado. O a nivel más general, si valoramos nuestra autonomía, el pensar por uno mismo, debemos dar más peso a nuestra posición en casos de desacuerdo entre pares.

Kelly, T. (2005). 'The Epistemic Significance of Disagreement'. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Vol. 1 (pp. 167–196). Oxford University Press.

Desacuerdos profundos

En otras ocasiones nos encontramos con gente que no solo discrepa en las conclusiones, sino en los criterios mismos para evaluar qué cuenta como una buena razón. Si yo creo en la evolución por selección natural porque acepto que la evidencia empírica es el estándar para conocer el mundo natural, mientras que tú rechazas la evolución porque consideras que la autoridad de tu texto sagrado es el criterio último de verdad, nuestro desacuerdo va más allá de interpretar datos diferentes.

Este es un desacuerdo más profundo: no compartimos los mismos principios sobre qué tipo de razones son válidas. Aquí no es que interpretemos la misma evidencia de forma distinta; discrepamos sobre los criterios de evidencia en primer lugar. ¿Qué debemos creer en estas situaciones en las que estamos en un desacuerdo profundo?

Desacuerdo entre pares: en un desacuerdo entre pares las dos personas comparten la misma evidencia, los mismos criterios de justificación y son igualmente razonables

Desacuerdo profundo: en un desacuerdo profundo las dos personas no comparten los mismos criterios de justificación

Matheson, J. (2021). Deep Disagreements and Rational Resolution. *Topoi* 40, 1025–1037.

Almagro, M. (2024). Deep conflicts and deep disagreements: Two ways in which a dispute can be deep. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 39 (1), 23–42.

El qué y el cómo en los desacuerdos

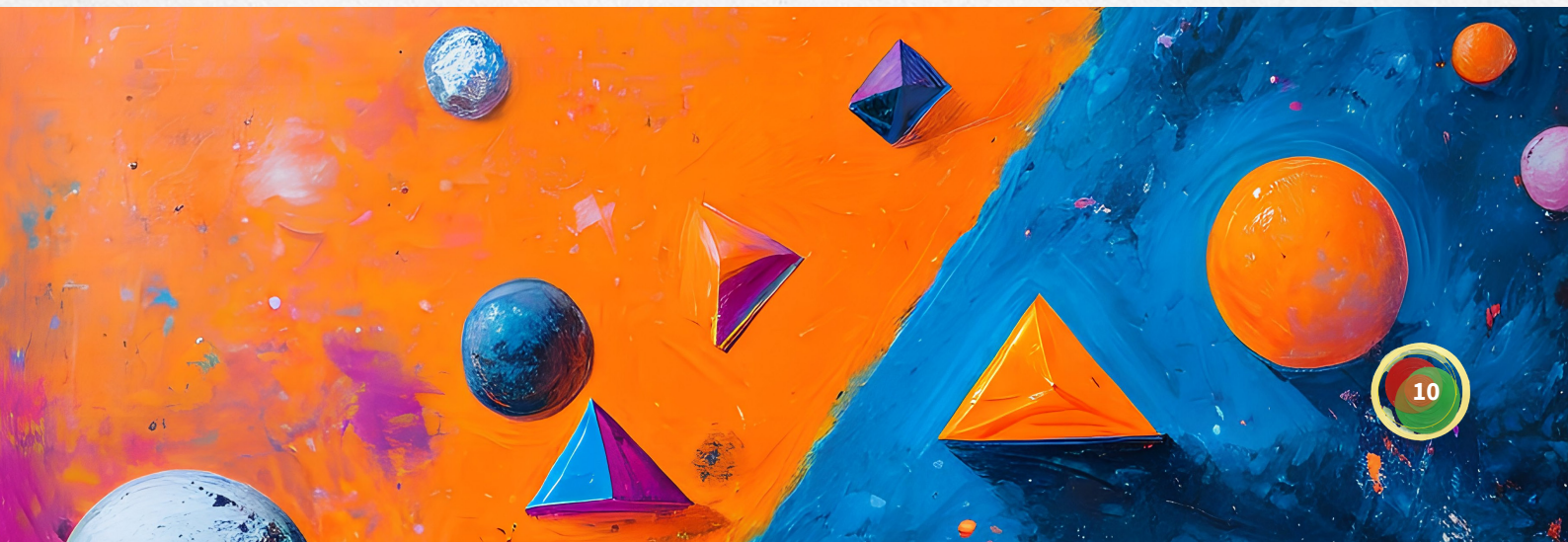
Además de la racionalidad que muestra cada hablante o de sus fuentes de justificación, los desacuerdos pueden separarse según una divisoria adicional, una que tiene que ver con el contenido y la actitud que mostramos hacia las creencias que expresamos en el desacuerdo.

En algunos desacuerdos hay grandes diferencias en lo que creemos acerca de un conjunto de asuntos o temas tan dispares como pagar impuestos, el aborto, los derechos reproductivos o la inmigración. En su versión menos benigna, la polarización ideológica implica que cada vez más personas se identifican con creencias ubicadas en los extremos del espectro político en relación con esos temas.

Otros desacuerdos se explican por la actitud de los agentes con relación a sus creencias. Esta actitud tiene que ver con nuestras disposiciones epistémicas implícitas o explícitas, así como con nuestras convicciones metaéticas. También con las emociones que expresamos en un desacuerdo. Cuando la polarización es de tipo afectivo (cuando se explica a partir de sentimientos que desplegamos pública y privadamente al estar en desacuerdo), puede ocurrir que el desacuerdo deje de ser una oportunidad para el diálogo y la reflexión y se convierta en una excusa donde lo que importa no es entender, sino vencer.

Miller, L. M. (2023). Polarizados. La política que nos divide, Madrid, Deusto.

Viciano, H. Hannikainen I. Gaitán, A. (2019). The dual nature of partisan prejudice: Morality and identity in a multiparty system. PLoS ONE 14(7)



Las distinciones anteriores apuntan a estructuras comunicativas y epistémicas ideales. En lo que sigue exploramos algunas preguntas acerca de cómo inciden los contextos digitales en esas estructuras.

¿Qué rasgos de las arquitecturas digitales favorecen que unos desacuerdos sean más frecuentes que otros? ¿Qué dispositivos y diseños favorecen las respuestas racionales frente a las dinámicas más perversas? ¿Cómo incide el contenido de los desacuerdos en las reacciones que mostramos cuando estamos inmersos en desacuerdos entre pares o en desacuerdos profundos? ¿Y cómo afectan las dinámicas de exclusión a la que están sujetos ciertos grupos a la resolución de estos desacuerdos?

Y a nivel más concreto: ¿Cómo se manifiestan estas dos caras de los desacuerdos en blogs y redes sociales (Facebook, X, TikTok)? ¿Cómo debemos organizar estos contextos para que la visibilización de desacuerdos no lleve aparejadas dinámicas más indeseables, como la polarización, el extremismo o las teorías de la conspiración?





La diversidad de opiniones no es siempre positiva

En un desacuerdo encontramos algo más que dos opiniones opuestas: debe darse además un intercambio de razones, puntos de vista, evidencia o argumentos. Pero incluso si asumimos esta complejidad, al pensar en los desacuerdos conviene empezar por el nivel más básico, esto es, el intercambio de opiniones opuestas. ¿Qué sabemos sobre el efecto de la diversidad de opiniones?

Se asume que el pluralismo de opiniones es casi siempre positivo. Dejando a un lado discursos que fomentan el odio o manifestaciones que desde una posición de autoridad podrían incitar a la violencia o tener consecuencias desastrosas, suele afirmarse que el libre intercambio de opiniones fomenta la calidad del debate, removiendo errores y afinando las posiciones en liza.

Mill, J. S. 'De la libertad de pensamiento y discusión' en Sobre la libertad, Madrid, Alianza.

Pero la evidencia científica ofrece un escenario diferente. La diversidad de puntos de vista puede fomentar la tolerancia y ser enriquecedora y beneficiosa para la deliberación, pero estos efectos positivos dependen, en gran medida, de las normas que las partes implicadas sigan al gestionar el desacuerdo.

Santoro, E., Broockman, D. E. (2022). The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments. *Sci. Adv.* 8, eabn5515 DOI:10.1126

Depende de las normas que se acepten

Cuando las normas que las posiciones abrazan favorecen la resolución cívica y razonada de los desacuerdos, es posible que se llegue a una mejora real en la comprensión, a veces de manera considerable. Pero pensemos en situaciones donde la polarización es intensa. Aquí las reglas del juego pueden ser muy diferentes. En estos contextos enconados, las partes pueden estar más motivadas a acentuar las diferencias que a resolver el desacuerdo. ¿Por qué? Algunas creencias no son acerca de si esta mañana va a llover o no para saber si tengo que coger un paraguas. Muchas creencias no tienen una función directamente instrumental, sino social: sirven para coordinarse con los miembros del grupo, para señalar que uno es un miembro aplicado del grupo en cuestión.

Funkhouser, E. (2022). A tribal mind: Beliefs that signal group identity or commitment. *Mind & Language*, 37(3), 444-464.

En contextos donde la señalización partidista es muy intensa, en lugar de buscar puntos en común o mejorar su comprensión, es habitual que muchos adopten una “mentalidad de soldado” y se guíen por la expectativa de no ceder ni mostrar acuerdo con el otro. Aquí, las normas que se siguen no son las de moderación ni las de resolución cívica del conflicto, sino las de reforzar la separación. En estos contextos resulta complicado acceder a los beneficios de los desacuerdos entre pares.

La resolución de desacuerdos entre iguales generalmente se favorece por el denominado “principio de caridad” que consiste en asumir la mayor racionalidad y actitud cooperativa posible en el que piensa diferente. Sin embargo, los desacuerdos en ambientes polarizados tienden a producirse en un trasfondo donde no rige el principio de caridad, ni la “mentalidad de explorador” orientada a resolver un problema, sino la caricaturización de los puntos de vista diversos y la señalización de lealtad hacia la bancada propia.

Galef, J. (2023). La mentalidad del explorador: Por qué algunas personas ven las cosas con claridad y otras no. Ediciones Paidós.

Viciano, H., Astobiza, A. M., Fasce, A., & Hannikainen, I. R. (2024). Scientifically Together, Politically Apart? *Sci & Educ*.

Diversidad de opiniones y moralización

Cuando las opiniones implicadas son morales la situación resulta aún más delicada. A lo largo de la historia, las diferencias morales han generado más rechazo social que otras formas de diversidad. Pensemos en el debate sobre la caza y su moralización: donde algunos asumen que se trata de una tradición valiosa para el control de la fauna, otros la describen como una práctica cruel y destructiva. En estos debates moralizados las diferencias de opinión suelen generar un gran rechazo entre quienes defienden y quienes condenan la práctica.

Las diferencias morales son como sabores fuertes. Si bien podemos tolerar una amplia gama de gustos o preferencias, los ‘sabores morales’ suelen despertar pasiones intensas: lo que unos encuentran aceptable, otros lo consideran repugnante. No tiene por qué ser así pero, al igual que alguien que evita ciertos restaurantes para no probar sabores que le resultan desagradables, muchas personas prefieren distanciarse de aquellos cuyas conductas o valores morales les resultan literalmente ‘indigeribles’.

Este rechazo no solo crea tensiones, sino que también puede fracturar las relaciones sociales y aumentar la distancia entre grupos. La distancia a su vez facilita la desconfianza y la caricatura del otro convirtiendo las diferencias morales en un terreno fértil para desacuerdos explosivos.

Haidt, J. Rosenberg, E. (2001). Differentiating diversities. Moral diversity is not like other kinds, *Journal of Applied Social Psychology*



Diversidad de opiniones y contextos digitales



*“No hay nostalgia peor que
añorar lo que no sucedió”*

Los efectos de la diversidad de opiniones también se han estudiado en contextos digitales. En lo que sigue atendemos a esa evidencia, pero, de entrada, deberíamos tener presente la famosa estrofa de la canción de Sabina que encabeza esta página.

Como afirma Luis Zaragoza (2024), historiador de medios y periodista en Radio Nacional de España: “decir, por ejemplo, que hoy debido a las redes sociales los individuos viven en burbujas en las que sólo se exponen a informaciones que concuerdan con sus opiniones e intereses podría inducir a pensar, por contraste, en un lector tipo que, pongamos por caso, en la España de los años treinta devorase cada día el monárquico ABC, el comunista Mundo Obrero, el católico «accidentalista» El Debate y el republicano liberal Ahora. Un lector idílico que, por supuesto, o nunca existió en el pasado, o fue francamente minoritario”.



Un poco de historia

En el primer lustro de la década del 2000 entró en escena la llamada web social, es decir, aquella en la que la interacción horizontal entre los usuarios cobra mayor protagonismo que el paradigma unidireccional y vertical entre productores de contenidos y usuarios. Internet dejó de ser una réplica digital del mundo analógico para convertirse en una cacofonía de blogs y primigenias redes sociales como MySpace. En esos primeros años, el ideal deliberativo estaba vivo y todavía resisten online varios intentos de mapear debates, como Debate Mapper, Debate Graph o Debate Map.

Levy, S. and Stone, B. (2006, 3 de abril). The new wisdom of the web. Newsweek.

Sin embargo, la web social pronto derivaría en un foro más participativo que deliberativo. Como si los creadores de Facebook hubieran intuido el dilema que planteaba Diana Mutz sobre la incompatibilidad entre deliberación y participación, la exitosa red social de Mark Zuckerberg se convertiría en un potentísimo canal de comercialización de publicidad incentivando la reacción emocional de los usuarios. La captura de la atención o ‘engagement’ tuvo un rédito económico inmediato que se potenció a través del interfaz, que permitirá a los navegantes declarar públicamente su preferencia por cierto tipo de contenidos, reenviándolos a su red y apostillándolos con comentarios propios.

Mutz, D. (2006). Hearing the other side: Deliberative versus participatory democracy. New York: Cambridge University Press.

Smith, B. (2023). Traffic: Genius, rivalry, and delusion in the billion-dollar race to go viral. New York: Penguin.

En el caso de Twitter, orientado más a la discusión pública que Facebook, el potencial deliberativo hace tiempo que sucumbió a las emociones y al sentimiento. Twitter se erigió casi desde sus inicios, más que en una plaza pública, como un ‘instrumento para detectar tendencias’. Es precisamente en la reacción pública al comportamiento de otras personas donde vemos el inicio del “espectáculo moral” que propician las redes sociales. Estas dinámicas también afectan a la calidad de los desacuerdos en contextos digitales. Aunque aquí atender a los detalles de cada contexto nos puede ayudar a separar casos positivos y negativos de señalización y postureo moral.

¿Cuánta diversidad hay en los contextos digitales?

Hemos visto que la diversidad de opiniones no implica necesariamente efectos beneficiosos. También sabemos que, cuando las opiniones son morales, los desacuerdos son más complicados de resolver. ¿Suced lo mismo en el contexto digital? ¿Cuánta diversidad de opiniones encontramos en este ámbito? ¿Y cómo afecta esa diversidad a la percepción de los desacuerdos en contextos digitales? Hay estudios que afirman que en redes sociales las personas tienen más posibilidades de exponerse a información no buscada («consumo incidental») que con las suscripciones a periódicos tradicionales.

Fletcher, R. (2021). The truth behind filter bubbles: Bursting some myths, Reuters Institute for the Study of Journalism, Oxford.

Algunos de estos estudios han demostrado que, en contraste con las ampliamente propuestas y testadas hipótesis de las «cámaras de eco» y «burbujas de filtro», las redes sociales favorecen la exposición a una mayor cantidad de opiniones con las que se está en desacuerdo.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F. and Volfovsky, A., 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), pp.9216-9221.

Barberá, Pablo. "How social media reduces mass political polarization. Evidence from Germany, Spain, and the US." *Job Market Paper*, New York University 46 (2014): 1-46.

Y parece que la exposición a ideas diversas o el acceso a puntos de vista opuestos no tiene porqué traducirse, necesariamente, en una mayor calidad de la deliberación pública. De hecho, algunos defienden que la diversidad de opiniones podría ser buscada por usuarios de redes sociales, aunque no siempre impulsada por principios de inclusión, sino más bien por la animosidad al otro.

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26).



Dos formas de estar aislado

El filósofo C. Thi Nguyen ha distinguido entre dos formas de aislamiento que nos pueden ayudar a entender lo que sucede en los entornos digitales. Nguyen diferencia una forma de aislamiento que se origina a partir de la estructura de conectividad de las redes, que, como sabemos, favorecen intereses y preferencias similares. En esas “burbujas epistémicas” accedemos a lo que nos gusta o nos resulta cercano a nivel ideológico o moral.

Pero además de las burbujas originadas por los filtros que nosotros creamos, Nguyen distingue otra estructura en las que el aislamiento se produce en torno al descrédito activo y la desconfianza hacia posiciones diferentes (“cámaras de eco”). Ahora no se trataría de ignorar otras fuentes, como sucede con las burbujas, sino estar dentro de comunidades o grupos en los que se fomenta la desconfianza hacia quienes opinan distinto y se exagera la fiabilidad de los que piensan como nosotros.

En las burbujas epistémicas las voces y la información divergente no llega; en las cámaras de eco la información llega, pero es desacreditada sistemáticamente desde posiciones muy radicales. Mientras que una burbuja epistémica puede ‘romperse’ aportando esa información que queda fuera como resultado de nuestro particular algoritmo, en el caso de la cámara de eco facilitar información opuesta o divergente ‘rebotará’, reforzando el punto de vista que se quiere corregir y sirviendo como herramienta para atacar al interlocutor.

Fletcher, R. (2021). The truth behind filter bubbles: Bursting some myths, Reuters Institute for the Study of Journalism, Oxford.

Nguyen C. T. (2020). Echo Chambers and Epistemic Bubbles, *Episteme*;17(2):141-161. doi:10.1017/epi.2018.32


¿Por qué Wikipedia o Reddit funcionan tan bien para los desacuerdos y X (antes Twitter) tan mal?

Un estudio sobre la sabiduría de las masas polarizadas muestra que, cuando las normas de resolución de desacuerdos son adecuadas, la diversidad ideológica puede enriquecer enormemente el debate, incluso en entornos polarizados. En plataformas como Wikipedia, donde rigen estas normas, los equipos con opiniones diversas suelen generar artículos de mayor calidad. Wikipedia también funciona como una red social. Sin embargo, poca gente reconoce Wikipedia como tal y, de hecho, esta imagen de desacuerdo productivo que proyecta no es la que generalmente tenemos de las redes sociales. ¿Por qué no percibimos el potencial productivo de estas plataformas para abordar los desacuerdos? ¿Por qué solemos verlas como un campo de batalla, un juego de suma cero, donde unos contendientes tratan de marcar puntos a expensas de otros que los pierden humillantemente?

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature human behaviour*, 3(4), 329-336.

Parte del problema puede explicarse por uno de los fenómenos más decisivos para entender nuestro mundo social: la auto-selección. Esta ocurre cuando personas con determinadas características tienden a agruparse en espacios o nichos que no reflejan de manera representativa a la población. Esto puede crear burbujas y simplemente generar la impresión distorsionada de que todo el mundo piensa o se comporta como ese grupo, que en realidad tiene características muy específicas.

Brundage, M., Little, A. T., & You, S. (2024). Selection neglect and political beliefs. *Annual Review of Political Science*, 27(1), 63-85.



Hay redes sociales que actúan como un acelerador de los desacuerdos morales y sus efectos más divisivos al potenciar la autoselección. La exposición a contenido divisivo desde grupos homogéneos puede distorsionar nuestras percepciones de los que piensan diferente. Las redes sociales amplifican estas distorsiones cuando exponen de manera desproporcionada a las personas a los extremos ideológicos.

Los usuarios más activos y comprometidos en las redes sociales suelen ser los más partidarios y extremistas, mientras que la mayoría de las personas, que son más moderadas, consumen el mismo tipo de noticias. Esto agrava el problema, generando la falsa impresión de que la polarización es más intensa de lo que realmente es.

Señalización cultural



Gran parte de los efectos negativos asociadas con los desacuerdos en contextos digitales dependen de dinámicas de señalización que se acentúan por los incentivos que regulan las interacciones en estos entornos - visibilidad, likes, seguidores, etc. Con ver ciertos perfiles de X o Facebook nos sentimos capaces de hacernos una idea del perfil político e ideológico del usuario, descontando su opinión cuando nos interesa, en el mejor de los casos, o recurriendo al insulto y a las malas formas en el peor de los escenarios.

La señalización de pertenencia a un grupo es una tendencia psicológica profunda, ampliamente estudiada. En contextos altamente polarizados, estas señales de identificación grupal - desde la bandera a los gustos musicales, pasando por las preferencias estéticas, los espacios y lugares o los hábitos de consumo - serán utilizadas por los grupos en disputa para facilitar su cohesión interna y para coordinar el descrédito de las posiciones opuestas.

Van der Does, T., Galesic, M., Dunivin, Z. O., & Smaldino, P. E. (2022). Strategic identity signaling in heterogeneous networks. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 119(10), 1–10.

Para algunos expertos, urge repensar el diseño de las redes sociales también en este sentido. Un modelo que minimice la incidencia de estas señales - que la mayor parte de las veces acompañan a perfiles anónimos - podría minimizar el conflicto. En esta línea se ha sugerido el fomento de espacios de interacción más neutrales (no anónimos o en los que el anonimato no se suplemente con señalizaciones culturales); otra posible intervención pasaría por limitar algunas métricas de interacción (limitar los ‘me gusta’ o los ‘retweets’) y reforzar la moderación de contenidos, por ejemplo recompensando en visibilidad a los que moderan contenidos con buenos criterios, como ya se hace en algunos foros de internet.

Bail, C. (2022). *Breaking the Social Media Prism. How to Make our Platforms Less Polarizing*, Princeton, Princeton University Press.

Transformando desacuerdos entre pares en desacuerdos profundos

La moralización de lo que antaño se podrían considerar preferencias personales (el consumo de tabaco, el uso del cinturón, reciclar, volar en avión o el uso de dispositivos electrónicos en la crianza) discurre en paralelo a la transición de la polarización ideológica a la afectiva: el rival político no es alguien que piensa diferente y con quien se puede llegar a un acuerdo, sino un enemigo moral cuyos valores suponen un riesgo existencial. El otro no propone otra visión del bien, sino que propone el mal, lo que dificulta cualquier tipo de deliberación.

Torcal, M. (2023). De votantes a hooligans: la polarización política en España. Madrid: Catarata.

Viciana, H., Hannikainen, I. R., & Rodríguez-Arias, D. (2021). Absolutely right and relatively good: Consequentialists see bioethical disagreement in a relativist light. *AJOB Empirical Bioethics*, 12(3), 190-205.

¿Qué nos dice todo esto sobre los desacuerdos entre pares y los desacuerdos profundos? Una hipótesis razonable apuntaría a cómo las redes sociales, además de animar el exhibicionismo moral, transforman los desacuerdos entre pares en desacuerdos profundos. Al llevar el desacuerdo desde el terreno de las preferencias o de las convenciones sociales al de la moral y las convicciones, todos los debates dejarían de lado sus aspectos técnicos e incluso el reconocimiento del interés personal o grupal, para convertirse en conflictos entre el bien y el mal. En el caso del uso de los aviones, por ejemplo, equivaldría al rechazo de este medio de transporte en cualquier circunstancia, sin entrar en las opciones disponibles u otras consideraciones.



¿Razones para la esperanza?

La evidencia de correlación entre polarización y ciertas redes sociales de internet parece clara. Pero no la relación de causalidad. Es decir, polarización y generalización de las redes sociales van de la mano, pero sería arriesgado aseverar que Facebook y X han incrementado la animosidad política de los ciudadanos. En otras palabras, la polarización y la expansión de las redes sociales podrían no ser causa y efecto directo, sino efectos paralelos de tendencias sociales más amplias, como el aumento de la desconfianza hacia las instituciones o la preferencia por contenidos emocionales y polarizantes.

¿Qué podemos afirmar con una mínima certeza? En su meta-revisión de la literatura sobre la relación entre redes sociales y democracia, Lorenz-Spreen et al. (2023) concluyen que existe una “asociación perjudicial entre medios digitales y polarización política, incluyendo alguna evidencia causal” (p. 7).

Lorenz-Spreen et al. (2023). A systematic review of world-wide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour* 7, 74–101.

El discurso catastrofista en torno a las burbujas epistémicas y las cámaras de eco ha tenido también mucha cobertura mediática y política. Pero solo un porcentaje muy reducido de ciudadanos (entre el 2% y el 5% en Europa y ligeramente más en EEUU) se encuentran en cámaras de eco informacionales. La mayoría de los ciudadanos acceden de forma frecuente a medios situados en un espacio ideológico lejano u opuesto - y esto es así incluso dentro de comunidades homogéneas insertas en plataformas sociales como X.

Levy, R. E. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3), 831-870.

¿Quién habita las cámaras de eco?

Se comienza a vislumbrar un perfil tentativo de usuario/a atrapado en cámaras de eco y ubicado en la intersección de varios rasgos: (i) muestra un elevado grado de compromiso con una determinada ideología política; (ii) tiene tendencia a descartar puntos de vista opuestos (y no meramente a seleccionar las opiniones afines); y (iii) expresa una elevada convicción en relación con opiniones morales y políticas. Estos perfiles tienen más probabilidad de estar dentro de redes homogéneas y fuertemente aisladas, redes que se estructuran en torno a temas o debates fuertemente divisivos. La buena noticia es que la gran mayoría de personas poseen perfiles de uso distintos al habitante de una cámara de eco.

Una paradoja

Los estudios empíricos sobre la exposición a información política en redes sociales revelan, por tanto, una intrigante paradoja: la tendencia mayoritaria entre la comunidad de usuarios es participar en redes sociales de composición heterogénea en las que la moderación es la tendencia mayoritaria – sobre todo aquellos interesados en la política, que muestran patrones de consumo mediático bastante variado. Sin embargo, una parte no despreciable del contenido político que el usuario mayoritario consume y comparte proviene de usuarios ideológicamente extremos, que son los responsables del contenido más viral en redes. Las consecuencias de este patrón están todavía por investigar y en este libro blanco nos contentamos con ponerlo sobre la mesa.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745

Tucker, J. A. et al. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19,).

Killian L. McLoughlin et al. (2024). Misinformation exploits outrage to spread online, *Science* 386, pp. 991-996

Gestionando el conflicto: ¿De enemigos a “frenemigos”?



Estudiosos de la polarización social y mediática han revelado que ciertos debates en X abren espacios de interacción y permiten intercambiar puntos de vista sobre temas tan conflictivos como puede ser la independencia de Cataluña. Esta exposición a informaciones sobre posturas políticas opuestas a las propias podría facilitar una mejora de las actitudes hacia el rival político.

Padró-Solanet, A., & Balcells, J. (2022). Media Diet and Polarisation: Evidence from Spain. *South European Society and Politics*, 27(1), 75–95.

Esta ‘despolarización’ podría darse solamente entre aquellos individuos de actitudes no muy extremas, es decir, que parten de posturas políticas iniciales moderadas respecto de su hostilidad hacia su oponente. Muy probablemente esta situación de disparidad moderada refleja desacuerdos entre pares, o incluso, relaciones «agonísticas», es decir, interacciones entre adversarios políticos legítimos.

Mouffe, C. (2000). *The Democratic Paradox*, London, Verso

Desde posturas críticas con el excesivo valor que otorgamos al consenso en la esfera pública, la deliberación sólo puede hacer avanzar la democracia mientras apoye perspectivas agonísticas – relación entre contrincantes legítimos– y no antagonistas –relación amigo-enemigo. Incluso desde esta perspectiva que celebra el conflicto de posturas, la solución pasaría por convertir a los enemigos políticos en «frenemies» u oponentes legítimos.

Alexander, J. Khosrokhavar, F. Stack, T. (eds). 2019 *Breaching the Civil Order: Radicalism and the Civil Sphere*, Cambridge University Press.

¿Podemos diseñar las redes sociales para acomodar dinámicas de desacuerdo?

Un factor clave para acelerar y acentuar los desacuerdos morales es la contagiosidad emocional, en particular, la propagación de emociones moralistas negativas como la indignación. Las plataformas sociales podrían actuar como ‘aceleradores’ de los desacuerdos morales al amplificar los sentimientos de rechazo y confrontación. Algunos estudios recientes revelan que la incivilidad entre figuras políticas ha aumentado en ciertas redes sociales como X, un reflejo de cómo las emociones negativas y moralistas se propagan rápidamente en estos espacios.

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), eabe5641.

Gervais, B. T., & Sanders, M. S. (2022). Incivility is Rising among Politicians on Twitter. *Social Psychological and Personality Science*, 13(4), 826-835.

El otro factor a tener en cuenta es que parte de lo que ocurre en la red no es más que un reflejo de lo que pasa fuera. Y es que los eventos políticos offline pueden hacer que las interacciones online sean aún más hostiles. Asimismo, la moderación de contenidos online puede reducir los eventos violentos y crímenes de odio fuera de internet. Todo esto sugiere que la influencia entre lo online y lo offline es bidireccional.

Hebbelstrup Rye Rasmussen, S., & Petersen, M. B. (2023). The event-driven nature of online political hostility: How offline political events make online interactions more hostile. *PNAS Nexus*, 2(11), pgad382

Jiménez Durán, R., Müller, K., & Schwarz, C. (2023). The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG. Available at SSRN 4230296.

¿La IA al rescate?




Además de cambios en la arquitectura de las redes sociales, la irrupción reciente de los modelos de lenguaje masivos (LLM) abre la posibilidad de intervenciones novedosas para minimizar el efecto negativo de los desacuerdos en los contextos digitales.

La potencia de los distintos chatbots de IA podría ayudar de entrada en la moderación de contenidos llevada a cabo por algunas plataformas. Otras intervenciones serían más sustantivas y podrían minimizar algunos de los problemas identificados en páginas anteriores. Por ejemplo, se ha ensayado con modelos de lenguaje que facilitan información factual en tiempo real y que ayudaría a minimizar dinámicas de desacuerdo; también se ha empezado a estudiar el efecto de chats que reformulan posiciones de consenso en una disputa, observándose una mayor disposición al acuerdo entre los participantes en esos foros de debate. Esto podría minimizar los efectos de los desacuerdos profundos. Aunque la IA no es ajena a los sesgos, seguramente su potencial puede ser positivo a la hora de reformular posiciones o enmarcar debates bien delimitados, en contextos específicos de desacuerdo.

Michael Henry Tessler et al. (2024). AI can help humans find common ground in democratic deliberation. *Science* 386,eadq2852(2024).DOI:10.1126/science.adq2852

L.P. Argyle, C.A. Bail, E.C. Busby, J.R. Gubler, T. Howe, C. Rytting, T. Sorensen, D. Wingate, Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale, *Proc. Natl. Acad. Sci. U.S.A.* 120 (41) e2311627120

Using GPT-4 for Content Moderation' (OpenAI, 15 August 2023)



“If you’ve been sexually harassed or assaulted write ‘me too’ as a reply to this tweet.” [«Si has sufrido acoso o agresión sexual escribe ‘yo también’ como respuesta a este tuit»]

Milano, A. [@Alyssa_Milano] (2017, 15 octubre) Twitter.

Rompiendo el silencio desde los contextos digitales

Tras destaparse en octubre de 2017 numerosas acusaciones de abusos sexuales contra el productor de cine Harvey Weinstein, la actriz estadounidense Alyssa Milano iniciaba con el mensaje que reproducimos en la página anterior un movimiento a gran escala de víctimas de abuso sexual que pusieron fin al silencio de muchos años. La publicación generó miles de respuestas y comentarios, e inspiró infinitud de publicaciones en redes sociales de mujeres y hombres de todo el mundo que, por primera vez, compartían sus historias personales. Sólo en las primeras 24 horas, el tuit alcanzó medio millón de retuits. Muchas celebridades respondieron, pero también lo hicieron un inmenso número mujeres que no eran conocidas: enfermeras, camareras, profesoras, ingenieras y estudiantes. Madres e hijas, hermanas y esposas.

Tambe, A. (2018). Reckoning with the Silences of #MeToo. *Feminist studies*, 44(1), 197-203.

#Metoo ejemplifica aspectos comunes a otros movimientos sociales que en los últimos años han canalizado desacuerdos activistas de manera exitosa en muchas partes del mundo (Líbano, Chile, Argentina, China, Francia...). Las expresiones de dichos desacuerdos tienen un carácter masivo. Se organizan de manera horizontal y fluida, la mayoría de las veces desde las redes sociales. Se expanden rápidamente a partir de un chispazo, un evento contingente, un post, una protesta minoritaria... Parten de demandas concretas para hacer un llamamiento más general a valores democráticos y de justicia social. Atraviesan las fronteras entre países hasta lograr apelar a una solidaridad transnacional.

Lester, L., & Cottle, S. (2022). Protests, publics and participation (still in an environmental age). In *The Routledge Handbook of Environment and Communication* (pp. 146-156). Routledge.

Activismo 2.0



En contraste con esta expresión del desacuerdo activista y digital, el activismo tradicional encuentra más obstáculos para generar espacios de oportunidad. En primer lugar, porque las propias condiciones de opresión dificultan la organización alrededor de una causa que puede no gozar de un estatus de credibilidad, confiabilidad o competencia, considerándose estas condiciones obstáculos para la paridad epistémica. La creación de nuevos espacios y narrativas de expresión desde contextos digitales, particularmente a través del empleo del *hashtag* como elemento enmarcador del discurso que señala “la potencia y la unidad del movimiento” (Dobrin), posibilitó prácticas de resistencia contra dichos obstáculos epistémicos.

Dobrin, D. (2020). The hashtag in digital activism: A cultural revolution. *Journal of Cultural Analysis and Social Change*, 5(1), 1-14.

Medina, J. (2019). “Racial violence, emotional friction, and Epistemic activism”. *Angelaki* 24(4), 22-37.

En segundo lugar, para operar de manera eficaz como movimiento político, el activismo generado desde espacios tradicionales (una plaza pública, un espacio de discusión), si bien no constituye un modo deliberativo de participación, se ve atado a la necesidad de generar una cierta interlocución con la sociedad o el grupo al que interpela. Por el contrario, los desacuerdos activistas online pueden estructurarse más allá de contextos deliberativos e incluso confrontativos con otros grupos o la sociedad mayoritaria. Pueden validarse como comunidades en sí mismas, generando afectos e identidades políticas comunes y en las que es suficiente el valor de expresar desacuerdo para señalar resistencia.

Frost-Arnold, K. (2023). *Who Should We Be Online?: A Social Epistemology for the Internet*, Oxford University Press.

Desacuerdos en contextos digitales y retóricas de la conspiración

Los desacuerdos morales no se producen en el vacío. Acabamos de ver que muchos desacuerdos se enmarcan en dinámicas de opresión y silenciamiento. Pero hay más dinámicas a tener en cuenta, algunas con un componente cultural muy marcado. En este sentido, las retóricas y narrativas de la conspiración han jugado un papel central en la percepción de desacuerdos en los contextos digitales.

En España las redes sociales han presenciado el negacionismo respecto a fenómenos como la existencia del COVID-19 o las enormes nevadas durante la borrasca Filomena de enero de 2021 —con algunas personas tratando de probar que la nieve no era real—. Las consecuencias prácticas del negacionismo sobre estos fenómenos son inmediatas y potencialmente catastróficas: rechazo o incluso sabotaje de las campañas de vacunación, actos de violencia, acoso a víctimas de desastres, etc.

Los estudios psicológicos y sociológicos en esta materia presentan hallazgos destacables. Lewandowsky, Oberauer y Gignac (2013) muestran que las teorías de la conspiración aparecen frecuentemente “empaquetadas”: la creencia en una teoría de la conspiración se correlaciona fuertemente con creencias en otras teorías de la conspiración, por lejanas que puedan parecer. Los estudios también han señalado que las identidades importan en la creencia de las teorías de la conspiración. En particular, Smallpage et al (2020) señalan que las personas son más propensas a creer en conspiraciones dirigidas contra su grupo, en particular si pertenecen a comunidades que ven como amenazadas, así muchas personas en África pudieron creer que el VIH o los medicamentos contra el VIH eran una creación para acabar con sus comunidades.

Lewandowsky, S., Oberauer, K., & Gignac, G. (2013). NASA Faked the Moon Landing-Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science. *Psychological science*, (24), 622-633.

Smallpage, S. M., Drochon, H., Uscinski, J. E., & Klostad, C. (2020). Who are the Conspiracy Theorists?: Demographics and conspiracy theories. In Butter, M., & Knight, P. (Eds.). *Routledge handbook of conspiracy theories* (pp. 263-277). Routledge.

¿Cómo minimizar la incidencia de las retóricas de la conspiración?

Para diseñar posibles intervenciones encaminadas a minimizar los efectos de la retórica de la conspiración, debemos tener muy presente que las teorías de la conspiración suelen “contaminar el pozo” y facilitan su difusión viral a base de sembrar desconfianza y cortocircuitar los mecanismos de pensamiento crítico. Esto significa que, una vez creída, la teoría de la conspiración proporciona razones para rechazar evidencia contra ellas. La naturaleza social de este tipo de creencias y su estatus como señalización de pertenencia grupal refuerzan este rasgo “memético” y apuntan a que las mejores intervenciones deberán enmarcarse a partir de este componente grupal o identitario.

Medidas como el *fact checking* (la creación de portales especializados en comprobar uno a uno la veracidad de ciertas informaciones) y el *flagging* (la señalización, por parte de la comunidad o la moderación de información engañosa) pueden servir para alertar a personas que no creen en la teoría conspiranoica. No obstante, pueden ser contraproductivas a la hora de argumentar con una persona ya convencida: al estar desacreditadas de partida, para una teoría de la conspiración estas medidas señalan justamente todo lo que la presunta conspiración quiere esconder. En términos de Krekó, pueden servir para la inmunización, pero no para la cura.

Debemos también tener presente el rol de los algoritmos que personalizan la información que recibe el usuario. Por tanto, si un usuario comienza a mostrar interés por teorías conspiranoicas, el algoritmo puede mostrarle cada vez más fuentes que lo reafirmen en su posición, en lugar de equilibrar ese punto de vista con visiones alternativas. Este fenómeno ilustra algo ya repetido en este libro blanco: en las intervenciones debemos atender no solo al contenido que se difunde, sino a los mecanismos (algoritmos) y a la arquitectura e incentivos de los entornos digitales. Se ha demostrado que los chatbots de IA también pueden jugar un papel eficaz aquí con intervenciones que pueden ser eficaces corrigiendo información conspirativa de manera interactiva, reduciendo su credibilidad y disminuyendo su difusión.

Boudry, M., Blancke, S., & Pigliucci, M. (2015). What makes weird beliefs thrive? The epidemiology of pseudoscience. *Philosophical Psychology*, 28(8), 1177-1198.

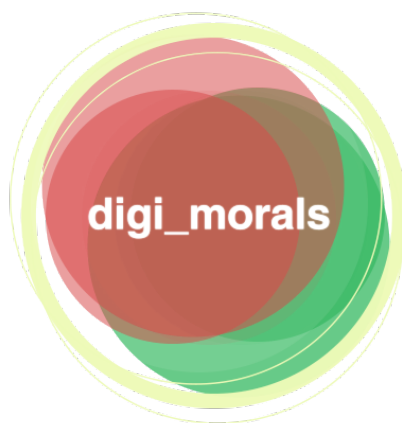
Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814.

Krekó, Péter. (2020) Countering conspiracy theories and misinformation. En: *Routledge Handbook of Conspiracy Theories*. Ed. por Butter, M; Knight, P. Londres: Routledge, 242-256.



Agradecemos la ayuda de los siguientes investigadores/as, que han participado en diferentes eventos dentro del marco del presente proyecto de investigación:

Uxía Carral, Simon Chauchard, Karen Frost-Arnold, Miriam Juan-Torres, Luis Miller, Peter Levine, Linda Skitka y Neri Marsili.



Este documento se ha elaborado durante el desarrollo del proyecto de investigación “Los desacuerdos morales en la esfera digital – dinámicas interactivas, micro-mecanismos y marcadores culturales”.

Este proyecto ha sido financiado por las Ayudas FBBVA a Proyectos de Investigación Científica, convocatoria 2021.

Fundación
BBVA