Daimon

Revista Internacional de Filosofía

Número 93. Septiembre-Diciembre 2024

DIVERSIDAD Y DELIBERACIÓN EN ENTORNOS DIGITALES

Editores:
Antonio Gaitán Torres, María Luengo Cruz y Gonzalo Velasco Arias

Democracia, deliberación y tolerancia en contextos digitales

Oportunidades y riesgos de los nuevos contextos digitales

Simposio sobre *Who Should We be Online* (OUP, 2023) de Karen Frost-Arnold

UNIVERSIDAD DE MURCIA
DEPARTAMENTO DE FILOSOFÍA

Daimon. ***Revista Internacional de Filosofía*****,** fundada en 1989, es una publicación cuatrimestral del Departamento de Filosofía de la Universidad de Murcia (España). Desde entonces, *Daimon* ha abierto un espacio filosófico de reflexión, análisis y crítica de problemas referidos principalmente al ser humano, en todas las dimensiones de su existencia.

Tiene como objetivo la publicación de investigaciones originales: publica por lo tanto trabajos que abordan, desde una perspectiva filosófica, las múltiples dimensiones o esferas de la existencia humana. *Daimon* es, pues, una revista que se dirige a investigadores, pero también a todo el que se interesa por el pensamiento filosófico en sentido amplio, desde la frontera de la ciencia hasta la de la literatura. En las páginas de *Daimon,* el especialista puede encontrar nuevos enfoques de un determinado problema o autor; el investigador, un espacio en el que publicar, contrastar o confirmar sus ideas; y el lector aficionado, artículos, revisiones críticas y reseñas de libros que pueden alimentar su curiosidad y ampliar su formación.

Daimon combina, en fin, el rigor académico con la originalidad de las investigaciones, sin olvidar la apertura y pluralidad necesarias en una publicación filosófica que quiere interesar también al lector ilustrado en general.

Daimon figura en el *European Reference Index for the Humanities* (ERIH) en la categoría ERIH PLUS *(Philosophy, 2016-01-11)*; sus artículos son registrados en las bases filosóficas de datos nacionales e internacionales siguientes: *Base ISOC - Filosofía. CINDOC* (España); *Dialnet* (España); *Francis, Philosophie. INIST. CNRS* (France); *Philosopher´s Index* (Bowling Green, OH, USA); *Repertoire Bibliographique de Philosophie* (Louvain, Belgique); *Ulrich´s Internacional Periodicals Directory* (New York, USA), *Scopus* (Editora Elsevier, Ámsterdam, Holanda), *Web of Science* (Clarivate Analitycs, Estados Unidos).

Daimon obtuvo por vez primera en 2014 el Certificado de Revista Excelente de la Fundación Española para la Ciencia y la Tecnología (FECYT, http://calidadrevistas.fecyt.es/Paginas/Home.aspx).

Edición electrónica: *www.um.es/daimon*

Daimon. *Revista Internacional de Filosofía*

Daimon. Revista Internacional de Filosofía

Publicación cuatrimestral. Número 93. Septiembre-Diciembre 2024

Monográfico sobre

Diversidad y deliberación en entornos digitales

Editores:
Antonio Gaitán Torres, María Luengo Cruz y Gonzalo Velasco Arias

Democracia, deliberación y tolerancia en contextos digitales

Oportunidades y riesgos de los nuevos contextos digitales

Simposio sobre *Who Should We be Online* (OUP, 2023) de Karen Frost-Arnold

UNIVERSIDAD DE MURCIA
DEPARTAMENTO DE FILOSOFÍA

Daimon. Revista Internacional de Filosofía

Publicación cuatrimestral. Número 93. Septiembre-Diciembre 2024

Administración: *Daimon* es una revista cuatrimestral, editada y distribuida por el Servicio de Publicaciones de la Universidad de Murcia. Apartado 4021. 30080 Murcia (España). Tfno.: 868883012. Fax: 868883414.

Redacción e intercambios: ver *Normas de publicación*, al final de la revista.

ISSN de la edición en papel: 1130-0507.
ISSN de la edición digital (disponible en http://revistas.um.es/daimon): 1989-4651.
Depósito legal: V 2459-1989.

Fundación BBVA Exclusivamente para la financiación de este número 93

Maquetación, diseño de cubierta: Compobell, S.L. Murcia.

Daimon. Revista Internacional de Filosofía
Publicación cuatrimestral. Número 93. Septiembre-Diciembre 2024

Monográfico sobre Diversidad y deliberación en entornos digitales

Artículos

Democracia, deliberación y tolerancia en contextos digitales

Oportunidades y riesgos de los nuevos contextos digitales

Simposio sobre *Who Should We be Online* (OUP, 2023) de Karen Frost-Arnold

Reseñas

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 5-15
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.622151

'Diversidad y deliberación en entornos digitales'

'Diversity and deliberation in digital contexts'

ANTONIO GAITÁN TORRES (UC3M)[*]
MARÍA LUENGO CRUZ (UC3M)[**]
GONZALO VELASCO ARIAS (UC3M)[***]

Abstract: This special issue aims to draw attention to the importance, opportunities and risks of deliberation in digital contexts. Contributions to this issue are not intended to be a state-of-art on this vast subject, but rather to open a window that portrays, like a still photograph, a plural set of approaches, problems, categories and concepts that tackle, in one way or another, on the general theme mentioned above. The window we are opening contains three views, those offered by Philosophy, Communication and Political Theory. In this brief introduction, we have articulated these three perspectives. The rationale of this special issue arises from the belief that the best approach to understanding the role and impact of deliberation in digital contexts is to take an interdisciplinary approach. Only from this 'hybrid' perspective (similar to the very nature of the mediations we face on a daily basis in digital environments) will we be able to offer diagnoses that are sensitive to the complexity of deliberation online as well as to adequately guide recommendations and interventions aimed at improving the quality of deliberation in such digital scenarios.

El objetivo de este número monográfico es llamar la atención sobre la importancia, las oportunidades y los riesgos de la deliberación en los nuevos y diversos contextos digitales. Lo que el lector tiene entre manos no ofrece un estado de la cuestión en torno a este vastísimo tema, pero sí quiere abrir una ventana que retrata, a modo de foto fija, un conjunto plural de enfoques, problemas, categorías y conceptos que tocan, de un modo u otro, el tema general que apuntábamos arriba. Ahora bien, ¿quien o quienes miran desde esa ventana?

---

* Datos biográficos disponibles en la p. 179.

** **María Luengo Cruz** (mluengo@hum.uc3m.es) es Profesora Titular de Periodismo del Departamento de Comunicación de la Universidad Carlos III de Madrid y Faculty Fellow del Yale Center for Cultural Sociology de Yale University. Lidera la Acción COST "Redressing Radical Polarisation: Strengthening European Civil Spheres facing Iliberal Digital Media (DepolarisingEU)" financiada por la Comisión Europea en el marco Horizonte Europa. Forma parte del equipo de investigación del proyecto "Desacuerdos morales en la esfera digital: Dinámicas interactivas, micro-mecanismos y marcadores culturales (digi_morals), financiado por la Fundación BBVA. Su trabajo analiza el rol de los medios de comunicación en la esfera civil desde la perspectiva de la sociología cultural y la teoría performativa. Entre sus libros más recientes destacan *The Crisis of Journalism Reconsidered: Democratic Culture, Professional Codes, Digital Future* (coeditado con Alexander y Breese, Cambridge University Press, 2016) y *News Media Innovation Reconsidered* (coeditado con Susana Herrera Damas, Wiley, 2021). Ha publicado en *American Journal of Cultural Sociology*, *European Journal of Communication*, *Media, Culture & Society*, *Journalism* y *Journalism Studies*.

*** Datos biográficos disponibles en la p. 169.

En nuestro caso, la ventana que abrimos contiene tres miradas, las que ofrecen la Filosofía, la Comunicación y la Teoría Política.

En esta breve introducción articulamos de forma preliminar las tres miradas desde las que hemos compuesto este número monográfico. Sin renunciar a la peculiaridad y al valor de cada uno de estos ámbitos, este monográfico surge de la convicción que el mejor enfoque para entender las variadas funciones de la deliberación en los contextos digitales, así como su incidencia en numerosas dinámicas, pasa por ejercitar una mirada interdisciplinar. Solo a partir de esa visión híbrida (como la propia naturaleza de las mediaciones a las que nos enfrentamos cotidianamente en los entornos digitales) podremos ofrecer diagnósticos que sean sensibles a la complejidad desde la que se articula la deliberación en redes sociales, blogs, etc. Y solo desde ese diagnóstico interdisciplinar podremos guiar de forma adecuada aquellas recomendaciones e intervenciones encaminadas a mejorar la calidad deliberativa en esos entornos.

Entre el optimismo y el malestar

Quedan muy lejanos aquellos años en los que se auguraba el inmenso potencial de las redes sociales para crear espacios de interacción que ayudarían a potenciar la calidad deliberativa de la esfera pública (Castells 2001. Dahlgren 2005. Negroponte 1995. Sunstein 2001). A principios de la década de 2010, el éxito de movimientos sociales espontáneos (el 15M en España, Occupy Wall Street en EEUU, la primavera árabe en el Magreb) coincidió con algunos relevantes desarrollos teóricos que preveían que la generalización de las redes sociales podría posibilitar un tipo de vínculo y de agencia colectiva que complementaría o vendría a sustituir a las formas tradicionales de participación política (Owen 2015. Tufekci 2017). Aunque alrededor de ese primer lustro de la pasada década los nuevos medios digitales permitieron el nacimiento de un tipo de agencia política y mediática novedosa, a esa promesa inicial le ha seguido un periodo más pesimista.

La eclosión de este pesimismo ha estado sin duda ligado a eventos políticos concretos (Brexit, Trump, Bolsonaro, etc.), pero también al sobredimensionamiento de algunos estudios empíricos que parecían encajar en esta potencial narrativa pesimista. En esos estudios se ha acentuado ciertas dinámicas de segregación que amenazarían los ideales deliberativos mencionados arriba (limitando la exposición a perspectivas y argumentos diferentes, por ejemplo, o afectando a la calidad y transparencia de la justificación) y afectarían al potencial deliberativo de los contextos digitales (Parisier 2011).

En épocas más recientes, sin embargo, hemos empezado a contar con evidencia más precisa y fina sobre los mecanismos y la incidencia efectiva de esas dinámicas y procesos de segregación en redes y plataformas digitales (Barberá et al. 2015). Esta imagen más reciente, todavía por precisar, relativiza la imagen pesimista de las redes sociales esbozada arriba, abriendo una pequeña puerta para la esperanza en relación con su potencial deliberativo y su capacidad para potenciar los efectos positivos de los desacuerdos. Las redes sociales seguramente no son el edén democrático que suponíamos, pero tampoco el campo de minas que hemos venido asumiendo en épocas recientes. Algunos hallazgos relevantes en este sentido serían:

- 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 (Gentzkow & Shapiro 2011).
- La mayoría de los ciudadanos acceden de forma frecuente a medios situados en un espacio ideológico lejano u opuesto (Dahlgren 2019) - y esto es así incluso dentro de comunidades homogéneas insertas en plataformas sociales como Twitter (Barberá 2015). Las redes sociales y los algoritmos que regulan su uso podrían estar exponiendo a los ciudadanos a más pluralidad que la que encuentran en contextos cotidianos – familia, trabajo, etc. (Bashky et al. 2015)
- Entre el reducido número de personas ubicadas en cámaras de eco se comienza a vislumbrar un perfil tentativo de usuario, perfil ubicado en la intersección de varios rasgos: (i) el grado de compromiso con una determinada ideología política; (ii) la tendencia a descartar puntos de vista opuestos (y no meramente a seleccionar las opiniones afines) (Quassam 2020) y (iii) la convicción que se expresa en relación con diferentes controversias (Iyengar & Hahn 2009). Este usuario tiene más probabilidad de estar dentro de redes homogéneas y fuertemente aisladas que se estructuran en torno al debate de ciertos temas o tópicos fuertemente divisivos.

Por tanto, a pesar del pesimismo que se observa en algunos círculos académicos, mediáticos y políticos, el ideal deliberativo no parece estar sujeto a la amenaza de las cámaras de eco, que se suponía general y simétrica entre los dos grandes campos ideológicos, sino más bien a los efectos nocivos que puede tener para la deliberación pública la acción más o menos orquestada de diversos grupos de extremistas que se organizan en torno a ciertos temas o controversias morales (inmigración, cambio climático, vacunación, etc.). Uno de los efectos más claros de estos grupos es la expulsión del ámbito deliberativo de posiciones moderadas o híbridas (Bail 2021). Los estudios empíricos sobre la exposición a información política en redes sociales revelan, en suma, una intrigante paradoja: los usuarios participan en redes sociales de composición heterogénea en las que la moderación es la norma – sobre todo si están interesados en la política y tienen consumos mediáticos heterogéneos (Dubois y Blank 2018) –, aunque una parte no despreciable del contenido político que consumen y comparten es ideológicamente extremo (Tucker et al. 2018). Cualquier análisis de la deliberación en los contextos digitales debe tener presente la evidencia anterior, tratando de determinar puntos comunes y diferencias específicas entre los distintos contextos deliberativos.

En cualquier caso, y a pesar de los esfuerzos de la ciencia social por tranquilizarnos, un malestar más difuso se ha instalado en torno a los entornos digitales, uno que en cierto sentido es impermeable al tipo de evidencia que acabamos de apuntar. Ese malestar tiene seguramente raíces históricas profundas, así que conviene decir algo sobre las mismas. Y en este tema, como en muchos otros, la potencia teórica de Max Weber puede servir de guía para entender algunas de las preocupaciones que acechan nuestra modernidad tardía. De acuerdo con la célebre tesis de Weber, la modernidad se caracteriza por una progresiva "diferenciación de esferas" de actividad y de sentido. Según este diagnóstico, la racionalización implicaría que los principios y fines que guían cada esfera de actividad se distinguen y especializan, generando en ocasiones contradicciones entre los fines de cada una de ellas y los valores de una sociedad: así, por ejemplo, la racionalización de la ciencia implica, por

un lado, el uso del método científico para la construcción de la objetividad, mientras que la lógica del mercado implica, por otro lado, la maximización del beneficio y de la eficiencia; o consideremos cómo la racionalización de la política conlleva su transformación en una técnica específica para la obtención del poder, que está en tensión con los valores democráticos y éticos que fundamentan su actividad.

Si razonamos asumiendo este diagnóstico general (que, sin lugar a dudas, ha sido ya puesto en entredicho por la mercantilización de todas las esferas de la vida en el orden neoliberal), la deliberación política, la deliberación lógico argumentativa y la deliberación ética deberían haber seguido caminos paralelos en su evolución, destinados por tanto a no encontrarse. Sin embargo, la esfera digital se ha encargado de difuminar los límites y separaciones asociados a esa especialización entre esferas deliberativas. La generalización del uso de dispositivos inteligentes asocia necesariamente la noción de "ciudadanos" con la de "usuarios" y "consumidores digitales". Y en la esfera digital, sobre todo en el ámbito de las redes sociales, se diluye la distinción entre ocio y profesión, participación y consumo, objetividad y opinión, solemnidad y frivolidad. En primer lugar, porque en virtud a lo que Chadwick ha llamado "el espacio mediático híbrido", ya no resulta fácil distinguir entre las prácticas y producción de contenidos de medios tradicionales y nuevos medios; como tampoco entre las prácticas comunicativas de actores políticos, periodistas y ciudadanos (Chadwick 2017). En segundo lugar, porque la ambivalencia y la ausencia de regulaciones explícitas para la conversación digital permiten la difusión de prácticas y actitudes a través de una mímesis no siempre intencional detonada por la extensión de temas y problemas formales propios de la política o de la ciencia a espacios informales de socialización digital (Kotsonis 2020). Como resultado, la deliberación racional, que se basa en la premisa de que los participantes actúan de manera racional, evaluando los argumentos presentados de acuerdo con criterios de validez lógica, coherencia y evidencia empírica, se entrevera motivaciones emocionales y afectivas vinculadas a la pertenencia tribal y a la búsqueda de reconocimiento; la deliberación ética, que funda su legitimidad no solo en la corrección lógica de los argumentos sino en la consideración moral hacia quien los emite, se confunde con actos de habla expresivos que manifiestan más bien el estado de ánimo y opinión del sujeto respecto a un tema conflictivo; la deliberación política, por último, relega los principios de igual acceso a la palabra y la opinión por la lógica casi bélica de la competición.

En este contexto, la deliberación resulta especialmente problemática desde el punto de vista práctico. La deliberación se torna ineficaz porque los agentes implicados en un mismo desacuerdo no siempre lo perciben y enfrentan desde la misma racionalidad (unos lo entienden como un problema científico, otros como un reto de índole moral; unos entienden que un disenso es de naturaleza política, otros que es una cuestión de moralidad básica). El estudio de estos "desacuerdos cruzados" (Osorio, J. Villanueva, N. 2019), a su vez, genera problemas para su estudio experimental: ¿en función de qué criterio convenimos que es preferible aplicar los estándares de evaluación de la deliberación política, de la ética o de la lógica? ¿Es necesaria esa aclaración analítica antes de afrontar la resolución de dilemas deliberativos detonados por los desacuerdos manifestados en las redes sociales?

Tres miradas en torno a los entornos digitales

La percepción de las oportunidades de los entornos digitales para potenciar los efectos beneficiosos de la deliberación ha ido variando del optimismo irrefrenable a un pesimismo general que, como acabamos de ver, se matiza desde la ciencia social y desde el análisis de las dinámicas culturales e institucionales. Aunque el escenario que conforman los nuevos entornos digitales no sea tan negativo para la deliberación como algunos agoreros nos quieren hacer ver, ciertamente se ha instalado un malestar difuso que requiere más análisis, a ser posible desde distintos ámbitos. En este número monográfico hemos tratado de articular ese espacio de encuentro que aúne tres posibles miradas al fenómeno de la deliberación en los nuevos entornos digitales. En esta sección repasamos cada una de estas perspectivas analíticas, ofreciendo al lector/a un marco en el que ubicar algunas de las aportaciones que se incluyen en este número monográfico.

A la filosofía moral le ha costado reparar en la importancia que tiene el mundo digital en nuestra vida cotidiana (Véliz 2021). La tendencia inicial, que pasaba por aplicar de manera automática marcos normativos bien conocidos (utilitarismo, deontología, virtudes) a las cuestiones que iban surgiendo al hilo de la incorporación de las nuevas tecnologías en diferentes ámbitos, se demostró pronto insuficiente. Algunas de las dinámicas y fenómenos que han ido emergiendo en los nuevos contextos digitales configuran espacios y marcos de interacción enteramente novedosos, por lo que la ética ha tenido que abordar esas cuestiones sin asumir las herramientas teóricas tradicionales y tratando de entender primero la conducta efectiva de los agentes en esos nuevos contextos (Levy 2019. Frost-Arnold 2023).

En este sentido menos orientado teóricamente, se pueden señalar, sin ánimo de ser exhaustivos, una serie de temas que han generado interés en la reciente ética digital. De entrada, destaca una línea de trabajo sustantiva centrada en entender las obligaciones morales y las virtudes que operan en entornos informacionales segregados (Mason 2018. Talisse 2019). Frente a la ortodoxia inicial, que evaluaba en sentido negativo cualquier entorno informacional segregado, en la actualidad abundan enfoques más complejos, que tratan de determinar qué tipos de contextos segregados son moralmente perniciosos, distinguiendo esas estructuras informacionales de otras que podrían tener efectos beneficiosos (Furman 2022) - más sobre este punto en el párrafo siguiente, acentuando la vertiente política. Otro gran tema en la reciente ética digital, ligado en parte al interés que acabamos de esbozar, tiene que ver con las virtudes o hábitos morales que debe ejercitar el agente en los nuevos contextos digitales. Virtudes epistémicas clásicas como la humildad han recibido tratamientos detallados en épocas recientes, en contraposición con vicios como la soberbia o el dogmatismo (Cassam 2018). También se han incorporado al análisis ético virtudes que son enteramente propias de los contextos digitales. La autenticidad y el 'fisgoneo virtuoso', por ejemplo, constituyen dos focos centrales del libro que articula el simposio incluido en este número monográfico (Frost-Arnold 2023). Otro gran foco de interés lo constituye el estudio de la privacidad y los cambios que esta categoría moral debe sufrir para acomodar nuestras intuiciones sobre el uso de información personal en contextos digitales (Véliz 2021). Finalmente, uno de los ámbitos de investigación más sugerentes en la ética digital se ocupa de la ética de aquellos actos de habla que son específicos de

los nuevos contextos digitales. La ética del 'posteo' o del 'retweet' ha concitado trabajo muy sugerente, trabajo que explicita el potencial normativo de enfoques propios de la pragmática y la filosofía del lenguaje (Marsili 2021). Finalmente, uno de los grandes temas en épocas recientes dentro de la ética digital aborda las dinámicas de moralización en contextos digitales (Brady et al. 2020. Tosi. Warmcke 2020). Entender cómo funcionan nuestras disposiciones y emociones morales básicas (resultado de un proceso evolutivo largo a partir de entornos presenciales 'próximos') en los nuevos entornos digitales resulta fundamental para cualquier propuesta de intervención centrada en el fomento del civismo y la tolerancia (Van Vabel y Packer 2021).

En el ámbito de la filosofía política y de la teoría social crítica, la principal preocupación se refiere a la evaluación de las dinámicas de grupos desde el punto de vista de la preservación del espacio público democrático, que es entendido fundamentalmente desde la teoría habermasiana. En primer lugar, se debate acerca de cómo analizar y clasificar los efectos de segregación de la opinión y de la identidad auspiciados por las distintas arquitecturas algorítmicas de las redes sociales. Si bien hay un cierto consenso preliminar a la hora de considerar a las "cámaras de eco" y a las "burbujas epistémicas" como problemas para el buen funcionamiento de la esfera civil habermasiana o de la razón pública rawlsiana, se ha argumentado también que un cierto grado de polarización ideológica puede ser sinónimo de ensanchamiento de la oferta política y de inclusión representativa. En consonancia con esa línea de discusión, las teorías sociales críticas que ponen el foco en la inclusión democrática y en la reversión de las injusticias epistémicas que traban el acceso a la deliberación, han sugerido que la creación de subesferas puede estar funcionando como una ocasión para superar las barreras de entrada que impone el modelo de la transparencia comunicativa habermasiana (Fraser 1992. Habgood-Coote et al. 2024). Los críticos con la deliberación así entendida sostienen que la exigencia de participar en un modelo normativo de deliberación definido por quienes ostentan posiciones de privilegio, tiene como efecto la exclusión o, incluso, la auto-exclusión de grupos y subjetividades no normativas, cuyas demandas se expresan más bien en protestas por injusticias concretas que se manifiestan de modo informal (Sanders 1997). Desde este punto de vista, la generación de subgrupos de opiniones afines fomentada por las redes sociales favorece la cohesión de los discursos discriminados, la creación de un acervo de experiencias compartidas, así como la autoconfianza de agentes que en una esfera públicas más abierta no se atreverían a participar. En esa línea, los estudios de caso recogidos por Karen Frost-Arnold en el libro que es objeto de discusión en este monográfico reflejan situaciones en las que colectivos feministas, racializados o interseccionales generan una razón compartida libre de los miedos y de las condenas de otros grupos de poder, lo que a su vez permite que sujetos privilegiados accedan a sus opiniones y testimonios sin imponer las condiciones y barreras de entrada que sí operarían de demandar a estos grupos que interviniesen en una esfera pública compartida (Frost-Arnold 2023). En resumen, el debate en la filosofía política y en la teoría social crítica se centra en dilucidar si la fragmentación grupal que favorece la digitalización del espacio público es siempre negativa o puede traer ventajas, y si esa nueva organización del debate requiere un enfoque normativo distinto en relación a la inclusión y el agenciamiento de nuevos sujetos políticos.

En el ámbito de la comunicación, buena parte de las problemáticas giran en torno al binomio información-técnica que, por lo demás, siempre ha acompañado a los grandes cambios

tecnológicos que han afectado a los medios de comunicación: la imprenta, la radio, la televisión e internet. "Lo digital" (y sus entornos), sin embargo, no se define ya como un medio más, ni siquiera como un sistema de medios en sí, como inicialmente se hizo en el caso de internet (Moragas 2012). El mundo de lo digital y virtual ha significado un paso disruptivo respecto al sistema de medios tradicional y ha abierto espacios de comunicación en los que coexisten la comunicación interpersonal, grupal y masiva. Son espacios en los que una misma persona puede apoyar una campaña para derribar a un gobierno o conseguir una mejora en su vecindario, crear su "alter ego" para las redes sociales, pedir un consejo sobre salud, hacer un pedido, difundir fotos íntimas, enterarse de una noticia e incluso generarla. Las implicaciones de este mundo digital en la comunicación son evidentes, pero aún las estamos intentando comprender. De entrada, la comunicación en "los espacios de flujo" digitales (Castells 2009) circula por un número casi infinito de conexiones. Las redes sociales permiten establecer un sinfín de influencias mutuas entre personas y grupos que no comparten el mismo espacio físico. Estas auténticas estructuras sociales desligadas del espacio y también del tiempo se dieron ya gracias a otras tecnologías de la comunicación anteriores a la digital (Thompson 1995). Pero las tecnologías digitales han convertido lo cuantitativo en cualitativo transformando desde dentro un sistema de medios que aún se encuentra en proceso de transición.

Desde teorías recientes del "actor-red" (Latour 2007), estudiosos de la comunicación han aplaudido la democratización de la información online a la que cualquier "nodo" de la red puede sumar aportes, pero igualmente desde estas teorías se han abierto puertas a la disolución en red de la mediación de actores e informadores relevantes como, por ejemplo, las organizaciones periodísticas, las cuales se convertirían en un nodo más de una gran conversación online. Igualmente, la polarización mediática en bloques ideológicos de medios antagónicos parece hacerse más intensa en entornos digitales, si bien siempre ha estado más o menos presente según los contextos y sistemas de comunicación pública. Junto a la agudización de líneas mediáticas divisivas, se plantean otras cuestiones de fondo en torno al compromiso con colectivos silenciados y vulnerables, la desinformación, la verificación, la ecuanimidad, la inclusión o la diversidad en la comunicación online. A estas y muchas otras cuestiones, se suma ahora el factor de la Inteligencia Artificial Generativa. Además, por destacar un último tema en relación a este número monográfico, los estudiosos y practicantes de la comunicación se preguntan por la proliferación de los así llamados "medios digitales", "pseudo-medios", "medios alternativos", "medios iliberales" u otros muchos calificativos que indican el desconocimiento que todavía existe alrededor de este fenómeno, de las comunidades online y offline que genera, y de las repercusiones que estos medios pueden tener (o están teniendo ya) en la calidad democrática de nuestros debates virtuales y, más importante aún, en la salud de nuestras sociedades reales.

Este número monográfico

A continuación describimos de forma breve el contenido de este número monográfico:

En **'Deliberación en democracias digitales: ¿es factible el ideal de una ciudadanía competente?', Rubén Marciel** analiza los principales problemas que tiene que afrontar la deliberación en el contexto actual, uno caracterizado por la fragmentación, la creciente

pluralidad de voces y una evidente erosión de diversas fuentes de mediación y autoridad. Asumido ese marco, Marciel defiende que el idea de ciudadanía competente es todavía factible en los nuevos contextos digitales, siempre que estemos dispuestos a acometer algunas reformas institucionales. El artículo de Marciel es un ejemplo de cómo integrar algunos debates clásicos en filosofía política dentro de los nuevos marcos comunicativos, testando la plausibilidad de conceptos, categorías y marcos de análisis que vienen de largo.

En un espíritu similar, **Andrea Carriquiry** se ocupa del ideal habermasiano de 'esfera pública', así como de las tensiones y potencialidades que conlleva pensarlo al hilo de los diversos entornos digitales contemporáneos - **'Deliberación en entornos digitales y tolerancia: repensar la esfera pública digital, con Habermas y más allá de Habermas'.** Carriquiry propone repensar el ideal habermasiano asumiendo la fragmentación de la esfera pública y apelando al sentido de tolerancia recientemente articulado por Rainer Forst.

Keberson Bresolin, autor del artículo **'Absolute Freedom of Speech and Social Media: Deconstructing the Argument of Individual self-Realization'**, deconstruye el argumento sobre la libertad de expresión absoluta como presupuesto para la autorrealización individual (Scanlon 1972). Teorías entusiastas de la era digital vieron en las redes sociales y la comunicación online en general canales privilegiados para esta expresión ilimitada de opiniones y para un debate abierto y transparente. La multiplicación exponencial del derecho a entablar discusiones fuertes sobre cuestiones controvertidas y desde visiones opuestas contribuiría a la autonomía personal y a la capacidad política de la ciudadanía, fomentaría el espíritu democrático (Shirky 2011) y tendría efectos disruptivos en el discurso dominante (Loader y Mercea 2011). Bresolin desmiente esta lógica de las redes sociales como espacios democráticos garantes de las libertades personales. Lo hace desde los postulados habermasianos que definen una conversación racional, ecuánime y cívica, en línea con otras aportaciones de este número monográfico.

En **'Microtargeting político y vigilancia social masiva', Carlos Saura García** describe los distintos procedimientos de microtargeting político que se usan para influir de forma selectiva en las preferencias políticas y pasa revista a las distintas medidas que se pueden tomar para paliar los efectos negativos de esta práctica. Al hilo de esta labor descriptiva, que sirve para enmarcar otros debates que aparecen en este número monográfico, Saura detalla el tipo de erosión al que se enfrenta la participación democrática contemporánea con la proliferación de este tipo de prácticas.

El manuscrito **'*Uncommon ground* y pluralidad de actos de habla en polílogos online'**, firmado por Catarina Machioni Spagnol, toma como referencia las aportaciones recientes de Aakhus y Lewiński (2017; 2023) a la teoría argumentativa en situaciones complejas y la noción de "polílogo" propuesta por estos autores para identificar nuevos patrones de comunicación en entornos digitales. El término "polílogo", según lo define Machioni Spagnol, alude a una conversación que involucra a múltiples participantes desde múltiples posiciones argumentativas y en múltiples lugares. A través de este concepto, la autora propone la superación de concepciones argumentativas más tradicionales como la basada en el modelo dicotómico proponente-oponente para entender mejor nuestros desacuerdos online.

En el artículo "¿Es la inteligencia artificial **doxástica un igual epistémico?", Alberto Murcia** se pregunta si las condiciones de paridad epistémica exigidas a un humano son aplicables a una Inteligencia Artificial Doxástica (IAD). El autor demuestra que, si bien la

igualdad cognitiva y la igualdad probatoria sí pueden ser satisfechas por una IAD, la condición de la revelación completa no puede ser totalmente alcanzada.

Raúl Tabarés Gutiérrez –Plataformización, automatización y aceleración en los medios sociales– sintetiza las características inherentes a los "social media" (blogs, wikis, redes sociales) que, según argumenta, impiden el fomento de espacios para la deliberación online. De acuerdo con Tabarés, la plataformización y automatización limitan el potencial de la conversación en redes estableciendo términos de referencia o mecanismos de moderación que reducen la complejidad y diversidad sociocultural que aflora en entornos digitales. Por su parte, la aceleración dificulta un intercambio reflexivo y reposado de opiniones.

La parte final de este número monográfico la ocupa un simposio centrado en ***Who Should You Be Online?: A Social Epistemology for the Internet*** **(OUP, 2023), de Karen Frost-Arnold**, uno de los libros recientes más importantes sobre la epistemología, la ética y la política en los nuevos contextos digitales. En su libro, Karen Frost-Arnold explora, valiéndose de casos prácticos y de un conjunto de 'personajes arquetípicos' que habitan en los nuevos espacios digitales, las diferentes y variadas formas en las que el ideal de conocimiento objetivo se ve amenazado o directamente subvertido. Según Frost-Arnold, ese ideal debe incluir la perspectiva de aquellos grupos excluidos y oprimidos, posibilitando espacios donde la confianza pueda establecerse y donde estas perspectivas puedan expresar sus opiniones y experiencias.

Además de los prècis y las réplicas de Karen Frost-Arnold, los tres artículos y la reseña que componen el simposio exploran algunas de las cuestiones centrales en torno a esta propuesta teórica, que como vemos tiene una marcada orientación hacia el activismo. **Lola Medina** explora las tensiones en el mantenimiento de la memoria colectiva en contextos virtuales acentuando la importancia de evitar la sobre-representación de ciertas creencias y la ausencia de creencias verdaderas no proposicionales. **Gonzalo Velasco** se centra en el tratamiento que Frost-Arnold hace de las virtudes epistémicas, en su caso el *lurking* o 'fisgoneo', para argumentar a favor de una perspectiva más social y menos individualista. **Antonio Gaitán** se ocupa de la noción de comunidad epistémica, esbozando los tres sentidos a los que apela Frost-Arnold en su libro y proponiendo una lectura menos optimista del potencial de las comunidades epistémicas cerradas y organizadas en torno a marcadores de identidad. En nuestro simposio también se incluye una esclarecedora reseña de **Beatriz Jordá** del libro de Frost-Arnold, que puede servir como una introducción perfecta a los temas del simposio.

* * * *

Este número monográfico se ha coordinado en el marco de las actividades del proyecto de investigación 'Los desacuerdos morales en la esfera digital - dinámicas interactivas, micromecanismos y marcadores culturales', Fundación BBVA - Proyectos de Investigación Científica 2021. Este proyecto ha servido de paraguas para las actividades de un grupo de filósofos/as y de expertos/as en comunicación y ciencia política de la Universidad Carlos III de Madrid y de Universidad de Sevilla, interesados en la incidencia de la deliberación en dinámicas como la polarización, el extremismo, la expresión de identidades digitales, el activismo, la desinformación, etc. Agradecemos el apoyo de la Fundación BBVA durante todo el desarrollo del proyecto, especialmente en lo que respecta a la coordinación y financiación de este número monográfico.

Queremos agradecer además la ayuda de las siguientes personas, sin cuyo trabajo desinteresado este número no habría sido posible: Fernando Aguiar, Manuel Almagro, Daniel Barbarrusa, Fernando Broncano, Rogério Christofoletti, Uxía Carral, Iván de los Ríos, Alicia García, Teresa Gil-López, Manuel Heras-Escribano, Javier López, Lola Medina, Alba Montes, Anibal Monasterio, Gabriela Müggenburg, Jesús Navarro, Felipe Núñez, Jon Rueda, Francisco Seoane, Marcello Sierra, Jesús Vega y Astrid Wagner. Sin la ayuda y la paciencia de Salvador Rubio, editor de Daimon, este número no habría cumplido con los plazos de publicación.

Referencias

Aakhus, M., & Lewiński, M. (2017). Advancing Polylogical Analysis of Large-Scale Argumentation: Disagreement Management in the Fracking Controversy. *Argumentation*, *31*(1), 179-207.

Aakhus, M. & Lewiński, M. (2023). *Argumentation in Complex Communication. Managing Disagreement in a Polylogue*, New York, Cambridge University Press.

Bail, C. (2021). *Breaking the Social Media Prism. How to Make our Platforms Less Polarizing*, New York, Princeton University Press

Barberá, P et al. (2015). 'Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber', Psychological Science, 1-12.

Bakshy, E., Messing, S. & Adamic, L.A. (2015). "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239), 1130-1132.

Brady WJ, Gantman AP, Van Bavel JJ. (2020). 'Attentional capture helps explain why moral and emotional content go viral'. Journal of Experimental Psychology General, 149 (4), pp.746-756.

Cassam, Q. (2018). *Epistemic Vices*, Oxford, Oxford University Press.

Cassam, Q. (2020). *Extremism*, London, Routledge

Castells, M. (2009). *Comunicación y poder*, Madrid: Alianza.

Castells, M. (2001). *La Galaxia Internet,* Barcelona: Plaza y Janés.

Chadwick, A. (2017). *The Hybrid Media System. Politics and Power*, Oxford, Oxford University Press

Dahlgren, P. M. (2019). Selective exposure to public service news over thirty years: The role of ideological leaning, party support, and political interest. International Journal of Press/Politics, *24*(3), 293–314.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. Political communication, 22(2), 147-162.

De Moragas, M. (2012) (ed.). *La comunicación: de los orígenes a internet*, Gedisa, Barcelona.

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.

Fraser, N. (1992). "Rethinking the Public Sphere: A Contribution to the Critique of Actually Democracy", in Craig J Calhoun, Habermas And The Public Sphere. MIT Press.

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

Furman, K. (2022). 'Epistemic Bunkers', *Social Epistemology*, Vol. 37, 2, pp. 197-207.
Gentzkow, M., & Shapiro, J. M. (2011). 'Ideological segregation online and offline' Quarterly Journal of Economics, *126*(4), 1799–1839.
Habgood-Coote, J., Ashton, N. A. & El Kassar, N., (2024) "Receptive Publics", Ergo an Open Access Journal of Philosophy 11: 5. doi: https://doi.org/10.3998/ergo.5710
Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication, 59*(1), 19–39.
Kotsonis, A. (2020). Social media as inadvertent educators. Journal of Moral Education, 51(2), 155–168. https://doi.org/10.1080/03057240.2020.1838267
Latour, B. (2007). *Reassembling the social: An introduction to actor-network-theory*. Oup Oxford.
Levy, N. (2021). *Bad Beliefs*, Oxford, Oxford University Press.
Loader, B. D., y Mercea, D. (2011). "Networking Democracy? Social Media Innovations and Participatory Politics." *Information, Communication and Society,* 14 (6), pp.757-769.
Marsili, N. (2021). 'Retweeting: Its Linguistic and Epistemic Value', *Synthese*, 198, 10457-10483
Mason, L. (2018). *Uncivil Agreement*, New York, Princeton University Press.
Negroponte, Nicholas.(1995). *Being Digital*. United States: Alfred A. Knopf.
Osorio, J. Villanueva, N. (2019). 'Expressivism and Crossed Disagreements', Royal Institute of Philosophy Supplement, 86, pp. 111-132.
Owen, T. (2015). *Disruptive Power: The Crisis of the State in the Digital Age*, Oxford University Press, Oxford / Nueva York.
Parisier, E. (2011). *The Filter Bubble: What Is the Internet Is Hiding from You*. The Penguin Press.
Sanders, L. M. (1997). Against Deliberation. Political Theory, 25(3), 347-376. https://doi.org/10.1177/0090591797025003002
Scanlon, T. (1972). "A Theory of Freedom of Expression", Philosophy *& Public Affairs*", 1 (2), pp.204-226.
Shirky, C. (2011). "The Political Power of Social Media Technology, the Public Sphere, and Political Change", *Foreign Affairs*, 90 (1), pp.1-9.
Sunstein, C R. (2001). *Republic.Com*. Princeton: Princeton University Press.
Talisse, R. 2019. *Overdoing Democracy*, Oxford, Oxford University Press.
Thompson, J B. (1995). *The media and modernity: A social theory of the media*. Stanford University Press.
Tosi, J. Warmcke, B. (2020). *Moral Grandstanding*, Oxford, Oxford University Press.
Tufekci, Z. (2017). *Twitter and Tear Gas. The Power and Fragility of Networked Protest*, New Haven, Yale University Press.
Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*.
Van Bavel, J. Packer, D. 2021. *The Power of Us*, New York, Little Brown
Véliz, C. (2021). *Privacidad es poder. Datos, vigilancia y libertad en la era digital*, Madrid, Debate.

DEMOCRACIA, DELIBERACIÓN Y TOLERANCIA EN CONTEXTOS DIGITALES

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 19-35
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.611681

Deliberación en democracias digitales: ¿es factible el ideal de una ciudadanía competente?

Deliberation in digital democracies: Is the ideal of a competent citizenry feasible?

RUBÉN MARCIEL[*]

Resumen: En este trabajo defiendo que el ideal de una ciudadanía competente es viable incluso en los contextos adversos que ofrecen las sociedades digitales. Para ello, identifico cinco problemas que obstaculizan a la ciudadanía la adquisición de competencia política: el pluralismo, el problema del moderador, la dificultad para acceder a información relevante, la apatía política y los sesgos políticos. Aunque estos problemas se agudizan en las democracias digitales, muestro que existen mecanismos institucionales que permiten corregir y mitigar sus efectos perjudiciales sobre la competencia política. Por ello, concluyo, el ideal de ciudadanía competente es *prima facie* factible y podría lograrse

Abstract: In this paper, I argue that the ideal of a competent citizenry is feasible even in the adverse context digital societies offer. To do so, I identify five problems that hinder citizens' acquisition of political competence: pluralism, the chairman problem, difficulties in finding relevant information, political apathy, and political biases. Even though these problems worsen in digital democracies, I show that there are institutional mechanisms that allow for the correction and mitigation of the detrimental effects they might have on political competence. I thus conclude that the ideal of a competent citizenry is *prima facie* feasible and could be achieved through the

Recibido: 09/04/2024. Aceptado: 18/06/2024.

* Investigador postdoctoral en el grupo *Law & Philosophy*, Universidad Pompeu Fabra. Líneas de investigación: teoría deliberativa, republicanismo, libertad de expresión, derecho a la información y ética del periodismo. Correo electrónico: ruben.marciel@upf.edu

** Últimas publicaciones: (i) Marciel, R. (2023). On citizens' right to information: Justification and analysis of the democratic right to be well informed. *Journal of Political Philosophy*, *31*(3): 358-84. https://doi.org/10.1111/jopp.12298; (ii) Marciel, R., y Magaña, P. (2023). (Not So) Happy Cows: An Autonomy-Based Argument for Regulating Animal Industry Misleading Commercial Speech. *Journal of Applied Philosophy*, Early view, 1-18. https://doi.org/10.1111/japp.12702.

*** Este trabajo ha sido financiado por el Ministerio de Educación Cultura y Deporte (beca predoctoral FPU ref. FPU2015-07227) y por la Fundació Irla (Beca Postdoctoral Irla d'Anàlisi i Pensament Social 2023-2024). El trabajo se enmarca en el proyecto de investigación "Razón Pública Global: Derechos Humanos, Legitimidad Democrática y Cambio Demográfico" (PID2020-115041GB-100/AEI/10.13039/501100011033), financiado por la Agencia Estatal de Investigación.

****Agradezco a Iñigo González-Ricoy, Adrián Herranz, Pablo Magaña y José Luis Martí los comentarios a las versiones previas del texto.

mediante la implementación de medidas institucionales que fomenten la adquisición de competencia política por parte de la ciudadanía.
Palabras clave: democracia, competencia, ciudadanía, deliberación, factibilidad.

implementation of institutional measures that foster political competence among citizens.
Keywords: democracy, competence, citizenry, deliberation, feasibility.

1. Introducción: el dilema democrático

Para que la democracia funcione adecuadamente es necesario que la ciudadanía cumpla mínimamente con ciertos estándares de competencia política. La competencia política (o cívica) consta de dos elementos: conocimiento político, que se adquiere mediante el procesamiento adecuado de la información relevante, y las capacidades necesarias para aplicar ese conocimiento en la realización satisfactoria de las obligaciones cívicas, como votar o manifestarse (Marciel, 2022: 70-72). La posesión de un mínimo de competencia cívica por parte de la ciudadanía aseguraría que las decisiones democráticas son adecuadas o, al menos, que no son nefastas.

Dada la necesidad democrática de un ciudadanía competente, pueden adoptarse dos posturas distintas. La posición *antidemocrática* asume que el ideal de una ciudadanía competente es irrealizable y, por tanto, que la idea una democracia funcional es también irrealizable (véase, e.g., Posner, 2003). La posición *democrática* defiende que el ideal de ciudadanía competente es plausible y, por tanto, que la democracia también lo es (véase, e.g., Innerarity, 2020). Así, ambas posturas, la democrática y la antidemocrática, no discrepan tanto sobre si la democracia es un ideal de gobierno deseable, sino sobre si podemos esperar que en la práctica la ciudadanía sea mínimamente competente. Esta es, claro, una discusión tan vieja como la democracia misma, y se retrotrae al menos hasta la Grecia clásica (véase Rapeli, 2014: cap. 2). Sin embargo, dos fenómenos relativamente recientes han reformulado el marco del debate.

El primer fenómeno, particularmente intenso durante los siglos XIX y XX, es la universalización del sufragio. El sufragio universal es relevante para este debate porque la expectativa de competencia política se extiende de manera análoga los derechos de participación: cuantas más personas participen en la toma de decisiones, más personas *deberían* ser políticamente competentes. Así, en nuestras democracias, donde (casi) toda la población adulta disfruta de derechos de participación política, (casi) toda la población adulta *debería* ser mínimamente competente (Brown, 1996). Esta exigencia generalizada de un mínimo de competencia política complica la causa democrática.

El segundo fenómeno, mucho más reciente y abrupto, es la revolución digital. Hasta finales del siglo XX, solamente la élite que controlaba los escasos medios de comunicación existentes tenía la capacidad para producir y distribuir públicamente contenidos. Sin embargo, la revolución digital hizo que tanto la creación como la transmisión de contenidos fuese mucho más rápida, sencilla y barata. Así, por primera vez en la historia de la humanidad la capacidad de generar y compartir contenidos se universalizó y dejó de ser oligopolio de la élite. Evidentemente, esta revolución digital ha tenido muchos efectos positivos (Benkler, 2006), pero también ha generado problemas—como una crisis

económica del periodismo (McChesney y Nichols, 2010: cap. 1) o el aumento de la desinformación (Wagner, 2022) y la polarización (Persily y Tucker, 2020: esp. caps. 2-3). Estos problemas revelan que las tecnologías digitales no garantizan ni un mejor funcionamiento de la democracia ni una mayor facilidad para adquirir competencia cívica. De hecho hay quien teme que las tecnologías digitales estén destruyendo la democracia (Bartlett, 2018; Curran, Fenton y Freedman, 2013).

A la luz de estos dos fenómenos, resulta particularmente difícil que en las democracias digitales la ciudadanía adquiera la competencia política necesaria para hacerse cargo de sus responsabilidades cívicas. Parecemos estar condenados a ese «dilema democrático» (Lupia y McCubbins, 1998: 1, 12) que nos obligaría a escoger entre dos opciones igualmente indeseables. La primera opción sería renunciar al sufragio universal y optar por un sistema epistocrático en el que sólo participen las personas más competentes. Así, se preservaría (al menos en teoría) la calidad de las decisiones políticas a costa de sacrificar la democracia. La segunda opción consistiría en preservar la democracia asumiendo que gran parte de la ciudadanía es incompetente y, por tanto, que la democracia generará malas decisiones. El dilema democrático resulta incómodo para cualquier demócrata porque incluso la opción menos mala implicaría renunciar a la aspiración de que las decisiones políticas sean la expresión de una voluntad popular *ilustrada*.

Nótese, sin embargo, que el dilema descansa sobre una asunción cuestionable, sin la cual no se plantea la necesidad de elegir entre estas dos opciones. Esa asunción es que en las sociedades de masas la ciudadanía es incapaz de adquirir la competencia política necesaria para encargarse de sus obligaciones políticas o, dicho de otro modo, que el ideal de ciudadanía competente es implausible. En este artículo intento mostrar que esa premisa es falsa. Para ello defiendo que, a pesar de las muchas dificultades que enfrenta en las sociedades digitales, el ideal una ciudadanía competente sigue siendo factible y que, por tanto, el ideal de una democracia funcional también lo es.

Para ello, en las siguientes secciones repasaré cinco problemas que sugieren la implausibilidad del ideal de ciudadanía competente: el pluralismo (sec. 2), el llamado problema del moderador (sec. 3), las dificultades para acceder a información relevante (sec. 4), la apatía política (sec. 5) y los sesgos cognitivos (sec. 6). Argumentaré que, a pesar de que todos estos problemas se agudizan en las sociedades digitales, existen mecanismos institucionales que nos permiten mitigar y/o corregir sus efectos perjudiciales sobre la competencia cívica, preservando así el valor epistémico en la toma de decisiones democráticas. Por tanto, el dilema democrático es un *falso* dilema que oculta una tercera opción: implementar medidas institucionales que protejan y promuevan la competencia cívica. En consecuencia, concluiré, el ideal de ciudadanía competente es factible si se implementan las medidas institucionales adecuadas.

Es importante anotar que esta discusión no trata sobre si la ciudadanía *es aquí y ahora* competente, sino sobre si el ideal de ciudadanía competente es *factible*. Lo factible es un tipo peculiar de posibilidad: aquello que, dado conjunto de hechos fijos, podemos conseguir intencionalmente a través de nuestros actos (Guillery, 2021). Así, al defender que el ideal de ciudadanía competente es factible defiendo que, a pesar de ciertos hechos fijos —como los sesgos cognitivos, la existencia de tecnologías digitales, o la dimensión masiva de nuestras democracias—, podemos conseguir a través ciertos actos intencionales —fundamentalmente

la reforma institucional— que la ciudadanía adquiera fácilmente el nivel de competencia cívica necesaria para desempeñar adecuadamente sus obligaciones cívicas.

2. El problema del pluralismo

Una de las características de las sociedades liberales es el pluralismo, esto es, la adopción por parte de sus miembros de doctrinas morales, políticas, filosóficas y religiosas distintas y a menudo incompatibles entre sí. Este fenómeno, al que Rawls (1996, 36) denomina «el hecho social del pluralismo», es el resultado natural al que inevitablemente llegamos las personas cuando vivimos en libertad. A menos que queramos restringir la libertad, tendremos pues que lidiar con el pluralismo que esta conlleva.

Surge así el problema democrático del pluralismo: la dificultad para alcanzar acuerdos aceptables en democracias modernas en las que existe una gran diversidad de creencias morales, filosóficas y religiosas, a menudo enfrentadas entre sí. La aceptación del pluralismo supone un reto democrático porque parece difícil que la ciudadanía, estando tan dividida sobre cuestiones tan fundamentales pueda alcanzar acuerdos políticos. Parece que para alcanzar cualquier acuerdo político la ciudadanía tendría que resolver antes sus profundas discrepancias, y que por lo tanto cualquier decisión política, para estar realmente legitimada, requeriría un inmenso esfuerzo deliberativo por parte de la ciudadanía. Esta problemática es aún mayor en sociedades digitales, porque ahora la ciudadanía puede acceder a más puntos de vista que en las sociedades predigitales. Asumiendo que distintos sectores de la sociedad adoptarán diferentes doctrinas, cabe esperar que el elenco total de puntos de vista en las sociedades digitales sea mucho más amplio que en las sociedades predigitales. Es decir, que como en las sociedades digitales hay más pluralismo, cabe esperar más dificultades para alcanzar acuerdos políticos.

Sin embargo, y en contra de lo que pudiera parecer, el pluralismo no supone un grave problema para el ideal de ciudadanía competente.

En primer lugar, es cierto que en una sociedad plural todas las decisiones políticas deberían estar públicamente justificadas (Marciel Pariente, 2020). Esto requiere que cada ciudadana tenga razones concluyentes para aceptar cada decisión política, pero no requiere que todo el mundo acepte cada decisión política por exactamente las mismas razones. De hecho, no parece necesario —ni, quizá, adecuado— exigir que los distintos miembros de la sociedad tengan motivaciones idénticas (Vallier, 2011). Piénsese, por ejemplo, en una política fiscal redistributiva que para los miembros de una doctrina religiosa es aceptable porque, en su visión, la redistribución de riqueza materializa el deber religioso de caridad. Para otros miembros de esa sociedad, la misma política fiscal podría resultar aceptable por otros motivos, como la creencia de que promueve un uso más eficiente de los recursos o una mayor igualdad de oportunidades. El acuerdo legítimo seguirá siendo posible a pesar de las discrepancias siempre y cuando las partes puedan encontrar razones concluyentes para aceptar esa política fiscal.

En segundo lugar, tengamos en cuenta que para alcanzar un acuerdo político no es necesario ponerse de acuerdo sobre absolutamente todo, ni tampoco sobre las distintas doctrinas morales, filosóficas, políticas o metafísicas. A pesar de que haya desacuerdos irreconciliables

sobre esas cuestiones, dado un mínimo consenso en torno a valores fundamentales, como la igualdad y la libertad, podemos alcanzar acuerdos sobre lo que hacer colectivamente y funcionar así como sociedad democrática (Rawls, 1996: cap. 6; 1997). En el caso de la política fiscal, por ejemplo, no es necesario ponerse de acuerdo sobre qué dios es el verdadero ni sobre qué principios morales deberían guiar nuestras decisiones. Lo único que tenemos que acordar es qué política fiscal debe implementarse. La respuesta a esta cuestión no implica adoptar ninguna postura sobre muchas de las cuestiones en las que el acuerdo es difícil o imposible. De hecho, como acabo de anotar, es muy plausible que una misma respuesta a la cuestión por la política fiscal tenga motivaciones distintas e incluso incompatibles.

Por tanto, la deliberación democrática, a través de la cual tratamos de resolver controversias políticas, no debería aspirar a resolver todos los desacuerdos, ni tampoco a alcanzar un consenso motivacional. Debería aspirar, tan sólo, a alcanzar lo que Martí denomina un «consenso operativo, esto es, un consenso sobre la decisión a tomar o la acción a emprender» (Martí, 2017: 12). Y si no es necesario ponerse de acuerdo en todo, sino únicamente en lo que hay que hacer colectivamente, y si no es necesario que todo el mundo esté de acuerdo por los mismos motivos, entonces las exigencias que para la ciudadanía se derivan de su deber de deliberar se reducen drásticamente. En concreto, no es necesario que la ciudadanía se informe sobre —ni mucho menos que comprenda y acepte— las doctrinas de sus conciudadanas. Estrictamente, tan sólo es necesario que la ciudadanía se informe sobre aquello cuyo conocimiento resulta útil para tomar decisiones colectivas adecuadas. En el debate sobre la política fiscal (por seguir con el ejemplo), no parece necesario que la ciudadanía piense lo mismo, digamos, sobre el dogma de la Santísima Trinidad; basta con que acuerde qué política fiscal es mejor. Y esto es compatible con que unos defiendan esa política fiscal movidos por convicciones religiosas —por ejemplo, de caridad cristiana— al tiempo que otros la defienden por consideraciones sobre justicia redistributiva. Así, el pluralismo no implica la implausibilidad del ideal de ciudadanía competente.

3. El problema del moderador

El segundo problema al que se enfrenta el ideal de ciudadanía es lo que, siguiendo a De Jouvenel (1961), denominaré *el problema del moderador*. El problema del moderador muestra cómo, a partir de un número relativamente bajo de participantes, la deliberación asamblearia resulta materialmente imposible.[1]

Si el tiempo de la asamblea se limita y se divide equitativamente entre las participantes, entonces cada participante tendría un tiempo irrisorio para hacer su intervención —tiempo que tendería a cero conforme aumentase el número de participantes. Así, por ejemplo, asumiendo 5.400 participantes —aproximadamente el número de ciudadanos de la Atenas clásica— y fijando 3 horas para el desarrollo de la sesión, cada participante tendría sólo 2 segundos para hacer su intervención (De Jouvenel, 1961: 368). Por otro lado, si tratamos de evitar este resultado estableciendo un tiempo mínimo decente para cada intervención, enton-

1 Aunque tomo el término del artículo de De Jouvenel, preocupaciones similares aparecen en las obras de, por ejemplo, Aristóteles (2010: 273, 1326a), Rousseau (1998: III-4), Hamilton (en Hamilton, Alexander Madison y Jay, 2009: nº9), Lippmann (1993: 35) o Lafont (2020: 28).

ces las sesiones durarían un tiempo inasumible —tiempo que tendería a infinito conforme aumentase el número de participantes. Retomando el escenario ateniense de 5.400 personas y otorgando a cada participante 15 minutos para intervenir, cada sesión duraría 1.350 horas. Es decir, que asumiendo jornadas de 9 horas al día, la asamblea tendría que reuniese durante 150 días antes de tomar una sola decisión (De Jouvenel, 1961: 369).[2]

El problema del moderador es un desafío para el ideal de ciudadanía bien informada porque, si las ciudadanas no pueden escucharse unas a otras parece difícil que puedan acceder a la información que necesitan para entender los asuntos públicos. Y así parece difícil que estén en condiciones de tomar decisiones adecuadas. La imposibilidad para escuchar a todo el mundo se magnifica en democracias digitales, ya que los contenidos (opiniones, noticias, sucesos, etc.) se multiplican a un ritmo vertiginoso. A pesar de eso, y como trataré de mostrar ahora, el problema del moderador no es tan grave como pudiera parecer, fundamentalmente porque *no es necesario que todo el mundo sea escuchado por todo el mundo*.

Tradicionalmente, la teoría democrática asumía que la democracia requería que el pueblo se reuniera para deliberar en asamblea y tomar decisiones políticas (Green, 2010). Gracias a este proceso de discusión asamblearia, las ciudadanas adquirirían la información y el conocimiento necesarios para tomar luego decisiones adecuadas. Este modelo de democracia asamblearia es típico de la antigüedad, propio de pequeñas sociedades como las ciudades-estado griegas, las tribus vikingas, o las pequeñas repúblicas tardomedievales y renacentistas (Dahl, 1998: cap. 2). Sin embargo, las democracias contemporáneas no son, ni pretenden ser, asamblearias. Todas las democracias de masas adoptan mecanismos de representación política que establecen lo que podría denominarse una *división del trabajo político* (Manin, 1997).

El reparto del trabajo político evita la necesidad de asambleas para la inmensa mayoría de la población, ya que divide a las ciudadanas en dos grupos con responsabilidades distintas. El primer grupo está formado por las representantes políticas, que son asesoradas por profesionales y expertas en las distintas materias. Estas personas son quienes deben dedicar mucho tiempo y esfuerzo a reunirse y discutir en detalle sobre las posibles decisiones específicas, porque son las que se encargan de tomar las decisiones finales, redactando y aprobando las leyes. El otro grupo está formado por la mayoría de las ciudadanas, que se limitan a indicar cuáles son sus preferencias y, en menor medida, qué medios prefieren para que esas preferencias sean satisfechas (Christiano, 1996: cap. 6; Lafont, 2020: 178). Para ello, las ciudadanas sólo deben indicar la dirección general que debería seguir la sociedad, ya sea mediante el voto o mediante otras formas de participación política, como las manifestaciones y las protestas, vigilando que sus representantes se esfuerzan por cumplir el mandato recibido de la ciudadanía. Así, a través del reparto del trabajo político, las democracias modernas combinan la representación con la participación, y permiten que las ciudadanas deliberen en diferentes

2 El problema del moderador plantea la duda de cómo era posible que la democracia ateniense funcionara. Lo cierto es que las asambleas de la democracia ateniense eran viables porque sólo un pequeño grupo de ciudadanos asistía a ellas, y porque de éstos sólo unos pocos intervenían en las discusiones públicas. A pesar de esto, la deliberación asamblearia sigue resultando implausible en sociedades de masas como las nuestras, pues incluso aunque –tal y como ocurría en Atenas– sólo un pequeño porcentaje de ciudadanas interviniese, ese pequeño porcentaje ya supondría miles de personas.

foros distribuidos por la sociedad sin necesidad de que todas ellas confluyan en un único espacio (Parkinson y Mansbridge, 2012).

Y esto que ocurre en la práctica es aceptado de manera mayoritaria en la teoría democrática, no sólo por pragmatismo, sino también porque, si se implementan bien, los mecanismos representativos promueven que las representantes políticas sean especialmente competentes (Landa y Pevnick, 2020). Efectivamente, la abrumadora mayoría de la teoría democrática contemporánea ve la democracia como un sistema representativo en el que la labor de la ciudadanía *no* consiste en participar regularmente en asambleas, sino en elegir a sus representantes, darles instrucciones y vigilarlas. En definitiva, la teoría política contemporánea ha abandonado lo que Green (2010: 9) denomina el «paradigma vocal» de la democracia para pasar a un «paradigma ocular».[3] En este paradigma la imposibilidad de que las ciudadanas se reúnan en una única asamblea y discutan todas con todas no es ningún problema, puesto que existen mecanismos de delegación gracias a los cuales sólo unas pocas personas *representantes* tienen que efectivamente reunirse y deliberar.

La ciudadanía puede, por tanto, cumplir satisfactoriamente sus obligaciones sin necesidad de reunirse en una asamblea en la que cada ciudadana sea escuchada por todas y cada una de las demás ciudadanas. Así, y en contra de lo que sugiere el viejo ideal de democracia asamblearia, para que la ciudadanía se informe adecuadamente lo importante no es que se produzca una deliberación asamblearia, sino que todo el mundo se informe sobre los asuntos democráticamente relevantes. La información sobre estos asuntos puede representar un porcentaje irrisorio respecto a toda la información disponible en la esfera pública, por lo que es factible que todas y cada una de las ciudadanas accedan a esa información. A propósito de la libertad de expresión Meiklejohn decía que lo importante «no es que todo el mundo pueda hablar, sino que todo lo que merece la pena decirse sea dicho» (Meiklejohn, 1965: 26); sobre el ideal de ciudadanía competente podríamos decir, de manera análoga, que lo importante no es que todo el mundo sea escuchado, sino que se escuche todo lo que merece la pena ser escuchado.

4. Problemas para acceder a la información relevante

En la sección previa he defendido que la imposibilidad de deliberar en una única asamblea no es problemática porque para estar en condiciones de tomar decisiones adecuadas la ciudadanía no necesita conocer todo lo que se dice, sino tan solo lo que es democráticamente relevante. El tercer problema al que se enfrenta el ideal de ciudadanía bien informada es, precisamente, la dificultad para acceder a la información relevante desde el punto de vista democrático.[4]

En un sentido muy básico, esta dificultad tiene un carácter puramente físico. Las dimensiones de las democracias de masas hacen que las personas y los eventos sobre los que hay que informarse estén demasiado lejos como para poder acceder a ellos directamente. Como

3 Quizá la única excepción importante aquí sea la teoría populista, la cual defiende un modelo de democracia radical sin intermediarios (Urbinati, 2015). En la práctica, empero, esa aspiración es inevitablemente abandonada (Taggart, 2000: 100).

4 Sobre la noción de relevancia democrática véase Marciel (2022: 76-80; 2023: 371-374)

dice Lippmann, «[e]l mundo con el que tenemos que lidiar políticamente está fuera de alcance, fuera de la vista, fuera de la mente» (Lippmann, 1991 [1922]: 29). Por otro lado, carecemos de los recursos necesarios (atención, tiempo y motivación) para escanear individualmente el vasto mundo que nos rodea y seleccionar por nosotros mismos la información más relevante. Tal y como anota de nuevo Lippmann, semejante tarea resulta inasumible a nivel individual:

> Al formar nuestras opiniones públicas, no sólo tenemos que visualizar más espacio del que podemos ver con nuestros ojos y más tiempo del que podemos sentir, sino que tenemos que describir y juzgar más gente, más acciones, más cosas de las que jamás podríamos contar o imaginar vivamente (Lippmann, 1991 [1922]: 148).

Planteada así, la dificultad para acceder a información democráticamente relevante no supone un gran reto para el ideal de ciudadanía competente, ya que las democracias modernas cuentan con dos elementos íntimamente relacionados, uno tecnológico y otro social, con el que superarla. El elemento tecnológico son los medios de comunicación, que permiten acceder fácilmente a información sobre personas y eventos lejanos. El elemento social, que complementa al tecnológico, es una nueva división del trabajo, a la que siguiendo a Bohman (2000) llamaré «división comunicativa del trabajo» (cf. Page, 1996: 2-6).

Esta división *comunicativa* del trabajo complementa la división *política* entre representantes y ciudadanía. En este caso, el reparto de tareas asigna distintos roles en el uso de los medios de comunicación. Por un lado, estarían las *comunicadoras profesionales*, que adquirirían el rol activo de buscar, seleccionar y proveer de información útil a la ciudadanía. Suele entenderse que el rol de comunicadora profesional incluye a cualquiera que se dedique habitualmente a la comunicación en la esfera pública, incluyendo a periodistas, políticas, miembros del gobierno, activistas, expertas, investigadoras, publicistas, comentaristas, grupos de presión, e incluso miembros de *think tanks* (Downs, 1957: 226; Page, 1996: 106-8; Zaller, 1992: 6). Por otro lado, estaría la *ciudadanía*, que tendría el rol pasivo de audiencia y se informaría con los contenidos ofrecidos por estas comunicadoras sin necesidad de buscarla directamente (Bohman, 2000: 55; Page, 1996: 4-6; Downs, 1957: 225-26). Así, las comunicadoras profesionales actuarían como intermediarias entre grandes audiencias de ciudadanas y el vasto, lejano y complejo mundo que las rodea. Como dice Page, estas comunicadoras «recopilan, explican, debaten, y diseminan las mejores informaciones e ideas disponibles sobre políticas públicas de manera que sean accesibles para grandes audiencias de ciudadanas corrientes» (Page, 1996: 5; cf. Bohman, 2000.).

De este modo, la conjunción de tecnología y de una adecuada organización social reducirían enormemente los costes de acceder a la información democráticamente relevante y posibilitarían el ideal de ciudadanía bien informada en grandes democracias. Esta es, no obstante, una visión demasiado idílica, pues —a pesar de la división comunicativa del trabajo y del uso de medios de comunicación— existen dos dificultades que ponen en cuestión la posibilidad de que la ciudadanía acceda fácilmente a información democráticamente relevante.

4.1. Desorden informativo

La primera dificultad, a la que siguiendo a Wardle y Derakhshan (2017) me referiré como *desorden informativo*, afecta al elemento tecnológico, es decir, a los medios de comunicación.

Como anotaba más arriba, desde finales del siglo pasado nuestra capacidad para generar y transmitir información de forma rápida, sencilla y barata no ha dejado de crecer exponencialmente. Según cuenta Pariser (2011: 11), toda la comunicación humana desde el inicio de los tiempos hasta 2003 ocuparía unos 5 mil gigabytes; en 2011, la humanidad generaba esa cantidad de datos cada dos días. Si esto era así en 2011, podemos asumir que ahora producimos esa cantidad de información mucho más rápido. Evidentemente, nadie puede filtrar individualmente todos esos contenidos y seleccionar lo más relevante. De ahí que haya quien se refiera al problema informativo de las sociedades digitales como «sobrecarga informativa» (Bartlett, 2018; Helberger, 2011: 242) o «superabundancia» (Ramonet, 1998) y «exceso de información» (Innerarity, 2011: 19, 25). Sin embargo, la cantidad de información disponible no es en sí misma un problema; el problema aparece cuando las grandes cantidades de información dificultan el acceso a lo que realmente importa (Helberger, ibid.; Ramonet, 1998: 42, 53-54, 195). Esto es lo que ocurre en nuestras sociedades digitales, y por eso el término «desorden informativo» resulta más apropiado que los términos que únicamente enfatizan la cantidad de contenidos.

El desorden informativo es la situación, típica de las democracias digitales, en la que las ciudadanas no pueden encontrar la información verdaderamente relevante, no sólo porque la cantidad de información sea abrumadora, sino fundamentalmente porque carecen de mecanismos fiables para filtrar los contenidos y seleccionar los que son veraces y relevantes. Si la ciudadanía se enfrenta a un océano infinito de información cuya credibilidad y relevancia son inciertas, parece efectivamente imposible que entienda los asuntos públicos y, por tanto, que pueda hacerse cargo de sus responsabilidades cívicas. Así, el desorden informativo parece poner en cuestión la plausibilidad del ideal de ciudadanía competente.

No obstante, nótese que este desorden tiene una base eminentemente institucional. No se trata de que la ciudadanía esté inevitablemente incapacitada para acceder, procesar y entender la información que necesita. Se trata, más bien, de que el contexto social actual hace esa labor extremadamente complicada. Ante este contexto adverso, no tenemos por qué renunciar a la aspiración de una democracia funcional; podemos, en cambio, intentar generar instituciones menos adversas, instituciones que ayuden a las ciudadanas a encontrar la información democráticamente relevante a un coste asumible. Esto nos lleva al segundo elemento —la división comunicativa del trabajo— y la segunda dificultad, a la que me referiré como *el problema de la identificación*.

4.2. El problema de la identificación

Como anoté arriba, las democracias representativas establecen una división de roles entre comunicadoras profesionales, por un lado, y ciudadanía, por otro. En teoría, esta división de tareas debería servir para mitigar los impactos del desorden informativo, puesto que la divi-

sión comunicativa del trabajo permitiría a la ciudadanía encontrar fácilmente la información que importan sin tener que filtrar ella misma todos los contenidos disponibles.

Sin embargo, y por más que la mayoría de las comunicadoras profesionales se presenten como servidoras del interés público, muchas no son fuentes de información fiables, ya sea porque tienen intereses que les impiden prestar un servicio informativo de calidad, ya sea porque simplemente son incapaces de hacerlo (por falta de talento o de recursos económicos). Así, para la ciudadanía resulta muy difícil, si no imposible, identificar qué fuentes de información son fiables. En estas condiciones es probable que se produzca lo que Buchanan (2018: 519) denomina «confianza epistémica descolocada», es decir, que la ciudadanía confíe en fuentes de información que no son de fiar y que, por ello, termine mal informada o desinformada.[5]

Lo anterior sugiere que la división comunicativa del trabajo es incapaz de solventar el desorden informativo. No obstante, una vez más el problema de fondo no es atribuible (al menos no solamente) a defectos de la ciudadanía, sino más bien a deficiencias del contexto institucional. Y, una vez más, este contexto institucional puede modificarse para facilitar que la ciudadanía pueda encontrar fuentes de información fiable. Hay al menos dos estrategias útiles para este fin. Una estrategia consiste en aumentar el número de fuentes fiables o, al menos, la proporción de fuentes fiables respecto a las no fiables. En las sociedades digitales existen infinidad de medios y canales que ofrecen contenidos sin garantía de veracidad ni relevancia; en cambio, las fuentes de información relevante y veraz —esto es, las fuentes que ofrecen periodismo de calidad— son relativamente pocas. Promover la proliferación del periodismo de calidad podría reequilibrar el elenco de fuentes disponibles, facilitando así el acceso de la ciudadanía a la información que necesita para estar bien informada (Marciel, 2023: 377-380). Una segunda estrategia, complementaria a la primera, consiste en facilitar el acceso a *metainformación*, esto es, a contenidos que ayuden a valorar la fiabilidad de las fuentes de información (Herzog, 2023: cap. 9). Si, por ejemplo, la ciudadanía pudiera saber los vínculos económicos o políticos de un periódico, podría valorar mejor su grado de independencia y ajustar su credibilidad.

Ni puedo ni pretendo discutir en detalle ninguna de estas estrategias aquí. Mi objetivo es únicamente mostrar que el problema de la identificación, al igual que el desorden informativo, pueden mitigarse mediante un diseño institucional adecuado. Este diseño debería facilitar a la ciudadanía el acceso a contenidos relevantes y veraces, así como el acceso a y la identificación de fuentes fiables capaces de ofrecer esos contenidos. En la medida en que exista margen para mejorar el diseño institucional, podemos confiar en que la supuesta incompetencia de la ciudadanía es (al menos parcialmente) corregible y, por tanto, en que el ideal de ciudadanía competente sigue siendo *prima facie* factible.

5. Apatía política

El cuarto problema al que se enfrenta el ideal de ciudadanía competente es la apatía política, la cual podría definirse como «la abdicación libremente elegida de la política por parte de aquellas ciudadanas que carecen del gusto por la vida cívica» (Green, 2004: 746).

5 Sobre la distinción entre mala información y desinformación, véase Marciel (2022: 82).

Suele asumirse que la apatía política forma parte de concepción moderna de la libertad. Siguiendo el famoso discurso *De la libertad de los antiguos comparada con la de los modernos*, de Benjamin Constant (1989 [1819]), la antigua concepción de la libertad entendía que ser libre significaba participar en los asuntos públicos; en cambio, la libertad en su sentido moderno consistiría más bien en todo lo contrario, en poder desentenderse de los asuntos públicos y dedicarse a disfrutar despreocupadamente de la vida privada. En realidad, la apatía no es un rasgo exclusivo de la ciudadanía moderna. La democracia ateniense, por ejemplo, también presentaba altas tasas de desinterés por los asuntos públicos, ya que muchos ciudadanos no solían acudir a las asambleas (Green, 2004). Sin embargo, suele asumirse que el grado de implicación en los asuntos públicos es generalmente menor en las grandes democracias de masas que en las democracias clásicas. Según Schumpeter, por ejemplo, en las democracias de masas la apatía política podría explicarse por la percepción de que los asuntos públicos tienen poco o nulo efecto en nuestras vidas, a diferencia de las decisiones sobre cuestiones estrictamente privadas —como dónde vivir, qué estudiar, con quien casarse, qué dieta seguir o cómo trabajar— las cuales tienen un efecto mucho más visible en nuestras vidas (Schumpeter, 2003 [1944]: 258, 261). Downs, (1957: 214), por su parte, ha ofrecido otra famosa explicación, conocida como *teoría de la ignorancia racional*: en sociedades de masas la influencia que tiene cada voto individual sobre la decisión final es prácticamente nula; sabiendo esto, la ciudadana racional carece de incentivos para informarse, y por eso prefiere dedicar sus esfuerzos a otros asuntos.

La apatía conlleva altos niveles de ignorancia sobre los asuntos públicos, y por ello plantea un problema para el proyecto democrático: a fin de cuentas, una ciudadanía que ignora los detalles de los asuntos públicos no parece estar en condiciones de tomar decisiones políticas. Sin embargo, en realidad la apatía política —y la ignorancia sobre los asuntos públicos que esta conlleva— no tienen por qué implicar necesariamente el fracaso del proyecto democrático. Hay, al menos, dos estrategias para evitar este problema.

La primera, favorecida por teóricos elitistas como el propio Schumpeter (2003: 295; cf. Brennan, 2011), es *desincentivar* la participación de aquellas ciudadanas menos competentes. Si las ciudadanas ignorantes no participan en los asuntos públicos, entonces los posibles efectos adversos de la ignorancia ciudadana desaparecen. Esta estrategia resulta problemática, en primer lugar, porque los desincentivos probablemente afectarían a la ciudadanía no de manera proporcional a su falta de competencia, sino de manera proporcional a su falta de convicción política. Así, probablemente la estrategia desmovilizaría a las personas con mayor nivel de autocrítica y menor autoestima, y no tanto las personas más fanatizadas y radicalizadas, que probablemente serían más refractarías a la desmovilización. En segundo lugar, la mera idea de desincentivar la participación parece contradecir la aspiración democrática de que la mayoría de la ciudadanía se informe y participe en los asuntos públicos. Sólo un teórico del elitismo democrático como Schumpeter puede percibir la baja participación como un síntoma de que la democracia funciona bien. Para la mayoría de las demócratas este síntoma sería, más bien, fuente de preocupación. Y, si sabemos que las instituciones y el contexto social afectan a la motivación para informarse (Boudreau, 2009; Kuklinski et al., 2001), ¿por qué no optar por un diseño institucional y unas políticas públicas que, en lugar de alejar aún más a la ciudadanía de la política, faciliten y estimulen la adquisición de competencia política y la participación de calidad?

La segunda estrategia consiste, precisamente, en esto. Para ello, resulta imprescindible diseñar instituciones que exijan de la ciudadanía la realización de tareas en un número y de una complejidad asequible. Por otro lado, es necesario que el diseño institucional facilite la adquisición de la motivación y el conocimiento necesario para desempeñar esas tareas. De esta forma, se reduciría el nivel de competencia necesario para desempeñar el rol ciudadano —ya que las tareas asociadas al rol serían menores— y, al mismo tiempo, se facilitaría la adquisición de ese nivel mínimo de competencia. En cierta medida, cualquier democracia representativa sigue esta estrategia al establecer la división del trabajo político mencionada más arriba. Efectivamente, en nuestros sistemas democráticos representativos la ciudadanía no se encarga de las cuestiones más técnicas; su papel consiste, más bien, en deliberar y, sobre la base de esa deliberación, escoger y fiscalizar a sus representes políticos. Esto reduce notablemente el número y la complejidad de las tareas exigidas a la ciudadanía y, en consecuencia, los estándares de competencia política (Christiano, 2015).

El viejo ideal de una ciudadanía super implicada y exhaustivamente informada resulta, pues, innecesariamente exigente (Moe, 2020). Las ciudadanas pueden adquirir mucha de la información que necesitan mediante «pistas» o «atajos informativos», esto es, contenidos fácilmente asequibles que permiten inferir qué decisión es preferible o qué conclusión es correcta sin necesidad de conocer todos los detalles sobre el asunto (Gilens y Murakawa, 2002). Por supuesto, no todos los atajos son igualmente útiles, ya que algunos llevan a conclusiones erróneas. No obstante, si se emplean bien, estos instrumentos heurísticos permiten tomar decisiones competentemente incluso a pesar de carecer de conocimientos sofisticados, sirviéndose de lo que Popkin (1991: 7) ha denominado «racionalidad de baja información».

Todo esto sugiere que, incluso con cotas relativamente altas de apatía política, si implementamos una adecuada división del trabajo político —que asigne a la ciudadanía tareas más acotadas— y buenas instituciones informativas —que proporcionen pistas y atajos informativos útiles—, entonces es plausible que la gente se informe suficientemente bien sobre los asuntos públicos y adquiera la competencia necesaria para tomar decisiones políticas bien fundamentadas.

6. Sesgos políticos

El último gran problema al que se enfrenta el ideal de ciudadanía competente son los sesgos cognitivos. Los sesgos cognitivos son fallos persistentes en el procesamiento de la información que nos predisponen a adoptar sistemáticamente creencias erróneas, desviadas en una determinada dirección. Así, evitan que creamos lo que, a la luz de la evidencia disponible, deberíamos creer.

Aunque la psicología experimental ha documentado muchos sesgos, los dos más notables son los de confirmación y desconfirmación. Los *sesgos de confirmación* son procesos que incrementan la probabilidad de aceptar argumentos o evidencia a favor nuestras creencias previas. Estos sesgos hacen que tendamos a validar la información que favorece lo que pensamos y a aceptarla sin someterla al debido escrutinio. También nos hacen dedicar más tiempo y esfuerzo a la búsqueda de información que corrobora lo que ya creemos, en comparación con el tiempo y esfuerzo dedicados a buscar información contraria a nuestros puntos

de vista. Los *sesgos de desconfirmación* también sirven para proteger nuestras creencias previas, pero protegiéndolas contra evidencias que las contradigan. De forma inversa a los sesgos de confirmación, los sesgos de desconfirmación aumentan el escepticismo y la crítica contra la información que choca con nuestras ideas previas y nos llevan a dedicar más tiempo y esfuerzo a contradecir esa información.

Aunque los sesgos afectan a todo razonamiento humano, su impacto es mucho mayor en el denominado razonamiento motivado por objetivos *direccionales* (Kunda, 1990: 480) o *partidistas* (Taber y Lodge, 2006: 756). Este tipo de razonamiento se produce cuando el sujeto prefiere alguna de las conclusiones que podrían derivarse del razonamiento, con independencia de si esa conclusión es correcta. El otro tipo de razonamiento al que suele referirse la psicología es denominado razonamiento motivado por objetivos de *precisión*, el cual ocurre cuando el sujeto desea llegar a una conclusión correcta, independientemente cuál sea. En estos casos, a pesar del esfuerzo que implica razonar, solemos seguir un proceso cognitivo cuidadoso, considerando minuciosamente la evidencia relevante y realizando inferencias precisas. Todo ello minimiza el efecto de los sesgos y aumenta las probabilidades de llegar a la conclusión correcta, que en última instancia es lo que pretende el razonamiento motivado por objetivos de precisión (Kunda, 1990: 480–82, 495; Taber y Lodge, 2006: 756).

Claramente, la política es uno de esos asuntos en los que tenemos preferencia por llegar a ciertas conclusiones con independencia de su corrección. A fin de cuentas, la política determina los derechos, los recursos y el reconocimiento que obtenemos. Por ello, cuando discutimos sobre asuntos políticos los sesgos se acentúan. Así lo demuestra la psicología empírica, revelando, por ejemplo, que algunas personas se niegan a corregir creencias erróneas incluso ante evidencia empírica que las contradice (Lewandowsky, Ecker y Cook, 2017; Thorson, 2016). Lo que estos estudios muestran es que algunas ciudadanas adquieren sus opiniones políticas no mediante la consideración de la evidencia más creíble y los argumentos más convincentes, tal y como presupone el ideal de ciudadanía competente (Taber y Lodge, 2006: 755), sino guiándose por sus emociones y su identidad. Así, lo esperable es que la ciudadanía termine creyendo no aquello para cuya creencia tiene más motivos, sino aquello que, supuestamente, alguien de su grupo político o social debería creer.

Todo esto pone en cuestión la posibilidad de una ciudadanía competente. Y, sin embargo, hay dos motivos para seguir considerando plausible el ideal de una ciudadanía competente.

El primero es que los sesgos cognitivos no siempre nos llevan a conclusiones equivocadas. Las cuestiones políticas no suelen tener una única respuesta correcta, sino que a menudo hay varias soluciones adecuadas. Especialmente en condiciones de pluralismo, distintas personas o comunidades pueden legítimamente tener preferencias distintas, sin que una de esas preferencias sea la correcta. En estas situaciones, los sesgos cognitivos pueden servir para identificar a quienes comparten nuestros intereses y para ayudarnos a articularlos y defenderlos de forma más eficiente (Lepoutre, 2020). Así, los sesgos cognitivos pueden servir como un mecanismo de protección de las minorías frente a la colonización cultural de las mayorías. De hecho, los foros de discusión *no plurales* (o *no públicos*), en los que un grupo de personas con puntos de vista similar discute apelando a sus valores compartidos, tienen valor democrático justamente porque sirven para reconocer y articular mejor las propias demandas del grupo en cuestión (Curran, 2002: 239). Por supuesto, estos foros no plurales pueden ser peligrosos si sus participantes no se exponen también a puntos de vista

e información contraria a sus ideas previas —de ahí la necesidad de que también existan foros públicos y plurales (Sunstein, 2007). No obstante, aquí tan sólo quería anotar que los sesgos cognitivos no implican necesariamente razonamientos erróneos, y que por tanto su persistencia no conlleva automáticamente la inviabilidad del ideal de ciudadanía competente.

Un segundo motivo para confiar en la factibilidad del ideal es que la intensidad de los sesgos cognitivos varía en función de ciertos factores ambientales que podemos modular (Thaler y Sunstein, 2008). Así, por ejemplo, la discusión intragrupal en grupos homogéneos tiende a exacerbar los sesgos previos, radicalizando las posiciones iniciales e intensificando la polarización; en cambio, si se delibera en grupos heterogéneos y el contexto facilita y motiva la adquisición de información útil, las personas consideran más detenidamente los contenidos, reflexionando de manera más imparcial (Mercier y Landemore, 2012; Kuklinski et al., 2001). Esto sugiere que, aunque los sesgos cognitivos sean inevitables, podemos diseñar instituciones que mitiguen sus efectos y promuevan una consideración más cuidadosa de la información, facilitando así la adquisición de competencia política.

Constatar que el efecto de los sesgos cognitivos se puede modular a través del diseño institucional es importante porque el diseño institucional de nuestras sociedades digitalizadas —y en particular, el funcionamiento de los medios de comunicación y de las redes sociales—resulta bastante mejorable: muchos de los medios de comunicación, tanto tradicionales como digitales, a los que la ciudadanía es expuesta regularmente a menudo no promueven la mitigación de los sesgos cognitivos, sino que más bien los exageran para promover sus intereses económicos y/o partidistas. Podemos, por tanto, imaginar diseños institucionales alternativos en los que los medios de comunicación y las redes sociales no estimulen (al menos, no tanto) esos sesgos. Por supuesto, no creo que construir una mejor esfera pública sea fácil; pero tampoco parece una tarea imposible. Lo importante es que en un mundo no muy lejano al nuestro es posible tener una ciudadanía menos afectada por los sesgos cognitivos, y que por tanto la presencia actual de esos sesgos cognitivos no implica *prima facie* la imposibilidad del ideal de ciudadanía competente.

7. Conclusión

Comencé este artículo presentando un «dilema democrático» que nos forzaba a elegir entre (a) mantener la democracia a pesar de que los procedimientos democráticos generarían malas decisiones o (b) preservar la calidad de las decisiones políticas renunciando a la democracia. A lo largo de las siguientes secciones he argumentado que esta disyuntiva es un falso dilema, porque asume implícitamente una premisa cuestionable, a saber: que la ciudadanía es irremediablemente incompetente.

Para ello, he abordado cinco problemas que dificultan la adquisición de competencia política en las sociedades digitales y que parecen sugerir la inviabilidad del ideal de ciudadanía competente: el pluralismo normativo, el problema del moderador, las dificultades para acceder a información relevante, la apatía política y los sesgos cognitivos. Si mi argumentación es correcta, existen mecanismos institucionales que nos permiten mitigar y/o corregir los efectos perjudiciales de todos estos problemas. Esto sugiere que el ideal de ciudadanía competente es *prima facie* factible incluso en los contextos epistémicamente adversos que

ofrecen nuestras democracias digitales. Por tanto, el falso dilema democrático puede evitarse optando por una tercera opción oculta por su planteamiento dicotómico: (c) implementar medidas institucionales que protejan y promuevan la competencia cívica. Queda por ver, pues, qué medidas son esas, y cómo podrían implementarse.

Referencias

Aristóteles (2010). *Política*. Madrid: Alianza Editorial.

Bartlett, J. (2018). *The People Vs Tech: How the internet is killing democracy*. Ebury Press.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Bohman, J. (2000). The Division of Labor in Democratic Discourse: Media, Experts, and Deliberative Democracy, en S. Chambers y A. N. Costain (eds.), *Deliberation, Democracy, and the Media* (pp. 47-64). Oxford: Rowman & Littlefield.

Boudreau, C. (2009). Making Citizens Smart: When Do Institutions Improve Unsophisticated Citizens' Decisions? *Political Behavior*, *31*(2), 287–306.

Brennan, J. (2011). *The Ethics of Voting*. Princeton: Princeton University Press.

Brown, R. D. (1996). *The Strength of a People. The Idea of an Informed Citizenry in America, 1650-1870*. Chapel Hill: University of North Carolina Press.

Christiano, T. (1996). *The Rule of the Many: Fundamental Issues in Democratic Theory*. Boulder: Westview Press.

Christiano, T. (2015). Voter Ignorance Is Not Necessarily a Problem. *Critical Review*, *27*(3–4), 253-269. https://doi.org/10.1080/08913811.2015.1111669

Constant, B. (1989). *De la libertad de los antiguos comparada con la de los modernos*. Madrid: Centro de Estudios Constitucionales.

Curran, J. (2002). *Media and Power*. Oxfordshire: Taylor & Francis.

Curran, J., Fenton, N. y Freedman, D. (eds.) (2013). *Misunderstanding the Internet*. London: Routledge.

Dahl, R. A. (1998). *On democracy*. New Haven: Yale University Press.

De Jouvenel, B. (1961). The Chairman's Problem. *American Political Science Review*, *55*(2), 368-372.

Downs, A. (1957). *An Economic theory of democracy*. New York: Harper Collins.

Gilens, M., y Murakawa, N. (2002). Elite Cues and Political Decision Making. *Political Decision Making, Deliberation and Participation*, *6*, 15-49.

Green, J. E. (2004). Apathy: the democratic disease. *Philosophy & Social Criticism*, *30*(4-5), 745-768. https://doi.org/10.1177/0191453704045763

Green, J. E. (2010). *The eyes of the people: democracy in an age of spectatorship*. New York: Oxford University Press.

Guillery, D. (2021). The Concept of Feasibility: A Multivocal Account. *Res Publica*, *27*(3), 491-507. https://doi.org/10.1007/s11158-020-09497-7

Hamilton, Alexander Madison, J., y Jay, J. (2009). *El Federalista. Fondo de Cultura Económica*.

Helberger, N. (2011). Media Diversity from the User's Perspective: An Introduction. *Journal of Information Policy*, *1*, 241-245. https://doi.org/10.5325/jinfopoli.1.2011.0241

Herzog, L. (2023). *Citizen Knowledge. Markets, Experts, and the Infrastructure of Democracy*. New York: Oxford University press.

Innerarity, D. (2011). *La democracia del conocimiento. Por una sociedad inteligente*. Barcelona: Paidós.

Innerarity, D. (2020), *Una teoría de la democracia compleja. Gobernar en el siglo XXI*. Barcelona: Galaxia Gutenberg.

Kuklinski, J. H., Quirk, P. J., Jerit, J., y Rich, R. F. (2001). The Political Environment and Citizen Competence. *American Journal of Political Science*, *45*(2), 410-424.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*, 480-498.

Lafont, C. (2020). *Democracy without Shortcuts*. New York: Oxford University Press.

Landa, D. y Pevnick, R. (2020). Representative Democracy as Defensible Epistocracy. *American Political Science Review*, *114*(1): 1-13. https://doi.org/10.1017/S0003055419000509.

Lepoutre, M. (2020). Democratic Group Cognition. *Philosophy and Public Affairs*, *48*(1), 40-78. https://doi.org/10.1111/papa.12157.

Lewandowsky, S., Ecker, U. K. H., y Cook, J. (2017). Beyond Misinformation: Understanding and coping with the post-truth era. *Journal of Applied Research in Memory and Cognition*, *6*(4): 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008.

Lippmann, W. (1991). *Public opinion*. New Brunswick: Transaction Publishers.

Lippmann, W. (1993). *The Phantom Public. Politics*. New Brunswick: Transaction Publishers.

Lupia, A., y McCubbins, M. A. (1998). *The democratic dilemma: can citizens learn what they need to know?* Cambridge: Cambridge University Press.

Manin, B. (1997). *The Principles of representative government*. Cambridge: Cambridge Univesity Press.

Marciel Pariente, R. (2020). Why Not Extend Rawls' Public Reason Beyond Fundamental Issues? A Defence of the Broad-Scope View of Public Reason. *Teorema*, *39*(2): 105-125. https://www.jstor.org/stable/26977735

Marciel, R. (2022). Democracia, Desinformación y Conocimiento Político: Algunas Aclaraciones Conceptuales. *Dilemata*, *38*: 69-86. https://www.dilemata.net/revista/index.php/dilemata/article/view/412000475/807

Marciel, R. (2023). On Citizens' Right to Information: Justification and Analysis of the Democratic Right to Be Well Informed. *Journal of Political Philosophy*, *31*(3): 358-384. https://doi.org/10.1111/jopp.12298.

Martí, J. L. (2017). Pluralism and consensus in deliberative democracy. *Critical Review of International Social and Political Philosophy*, *20*(5), 556–579. https://doi.org/10.1080/13698230.2017.1328089

McChesney, R.W., y Nichols, J. (2010). *The Death and Life of American Journalism*. New York: Nation Books.

Meiklejohn, A. (1965), *Political Freedom: the Constitutional Powers of the People*. Oxford University Press.

Mercier, H. y Landemore, H. (2012). Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation. *Political Psychology*, *33*(2): 423-436. https://doi.org/10.1111/j.

Moe, H. (2020). Distributed readiness citizenship: A realistic, normative concept for citizens' public connection. *Communication Theory*, *30*(2), 205-225. https://doi.org/10.1093/CT/QTZ016

Page, B. I. (1996). *Who deliberates?: mass media in modern democracy*. Chicago: The University of Chicago Press.

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.

Parkinson, J., y Mansbridge, J. (eds.) (2012). *Deliberative Systems. Deliberative Democracy at the large Scale*. Cambridge: Cambridge University Press.

Persily, N. y Tucker, J.A. (eds.) (2020). *Social Media and Democracy. The State of the Field and Prospects for Reform*. Cambridge: Cambridge University Press.

Popkin, S. L. (1991). *The reasoning voter: communication and persuasion in presidential campaigns*. Chicago: University of Chicago Press.

Posner, R. A. (2003). *Law, pragmatism, and democracy*. Cambridge: Harvard University Press.

Ramonet, I. (1998). *La tiranía de la comunicación*. Madrid: Debate.

Rapeli, L. (2014). *The Conception of Citizen Knowledge in Democratic Theory*. Hampshire: Palgrave Macmillan.

Rawls, J. (1996). *Political Liberalism*. New York: Columbia University Press.

Rawls, J. (1997). The Idea of Public Reason Revisited. *University of Chicago Law Review*, *64*(3), 765-807. https://doi.org/10.2307/1600311

Rousseau, J.-J. (1998). *El contrato social*. Madrid: Tecnos.

Schumpeter, J. A. (2003). *Capitalism, Socialism and Democracy*. London: Routledge.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton: Princeton University Press.

Taber, C. S., y Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, *50*(3), 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Taggart, P. A. (2000). *Populism*. Buckingham: Open University Press.

Thaler, R. H., y Sunstein, C. R. (2008). *Nudge. Improving Decisions About Health, Wealth, and Happiness*. New Heaven: Yale University Press.

Thorson, E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, *33*(3), 460-480. https://doi.org/10.1080/10584609.2015.1102187

Urbinati, N. (2015). A Revolt Against Intermediary Bodies. *Constellations*, 22(4), 477-486. https://doi.org/10.1111/1467-8675.12188

Vallier, K. (2011). Convergence and Consensus in Public Reason. *Public Affairs Quarterly*, *25*(4): 261-79.

Wagner, A. (2022). Retos Filosóficos de Las Sociedades Digitales: Esbozo de Un Enfoque Sistémico. *Dilemata*, *38*: 13-29. https://www.dilemata.net/revista/index.php/dilemata/article/view/412000497/797

Wardle, C., y Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Strasbourg.

Zaller, J. R. (1992). *The Nature and origins of mass opinion*. New York: Cambridge University Press.

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 37-54
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.612101

Deliberación en entornos digitales y tolerancia: repensar la esfera pública digital, con Habermas y más allá de Habermas*,**

Deliberation in digital environments and toleration: rethinking the digital public sphere, with Habermas and beyond Habermas

*ANDREA CARRIQUIRY****

Resumen: En este artículo analizo la esfera pública digital *qua* esfera pública, incluyendo herramientas de quien acuñó dicha noción, Jürgen Habermas, y publicó en 2022 un libro donde trata la esfera pública digital. Analizo el fenómeno de la fragmentación —aspecto que puede darse asociado a la diversidad en entornos digitales—; examino la cuestión del signo positivo o negativo de la esfera pública digital para la democracia, y propongo, como herramienta teórica en línea con la deliberación, una conceptualización de la noción de tolerancia que abreva en el trabajo de Rainer Forst, lo articula con lo anterior y lo expande.
Palabras clave: esfera pública digital, democracia deliberativa, Jürgen Habermas, Rainer Forst, tolerancia.

Abstract: In this article I analyze the digital public sphere qua public sphere, including tools from the author who coined such notion, Jürgen Habermas, and published a book in 2022 that discusses the digital public sphere. I analyze the phenomenon of fragmentation —an aspect that can occur associated with diversity in digital environments—; I examine the question of the positive or negative sign of the digital public sphere for democracy, and I propose, as a theoretical tool in line with deliberation, a conceptualization of the notion of toleration that draws on the work of Rainer Forst, articulates it with the above and expands it.
Keywords: digital public sphere, deliberative democracy, Jürgen Habermas, Rainer Forst, toleration.

Recibido: 13/04/2024. Aceptado: 25/06/2024.

* Agradezco el trabajo de las revisiones anónimas y del equipo editorial, que ha contribuido a mejorar este artículo en diversos aspectos. La responsabilidad de los errores remanentes es, por supuesto, mía.

** Artículo realizado en el marco del Plan de Trabajo del régimen de Dedicación Total de la autora: «Teoría de la democracia deliberativa y esfera pública: conceptualizando los discursos de odio, "noticias falsas" y sociedad civil "incivilizada" en torno a los feminismos latinoamericanos contemporáneos», financiado por la Universidad de la República.

*** Doctora en Filosofía. Docente investigadora en régimen de Dedicación Total, Universidad de la República. Líneas de investigación: teoría de la democracia deliberativa, movimientos sociales y esfera pública digital. Ha dictado clases a nivel de grado y posgrado, y publicado en libros y revistas arbitradas dentro y fuera de Uruguay, sobre temas de filosofía política, estética y filosofía nacional. Publicaciones recientes: «Pensar con Habermas, después de Habermas: el rol de la prensa en la esfera pública (digital)», *Sistema, 263*, 49-64; «Jürgen Habermas y lo privado vuelto al público, en la esfera pública original y en la esfera pública digital», *Ideas y Valores, 71*(180), 123-146. Correo electrónico: andrea.carriquiry@fhce.edu.uy

Introducción

Este artículo es resultado de una investigación más amplia; a los efectos de este monográfico me enfoco en dos aspectos que pueden relacionarse con la diversidad y la deliberación respectivamente: el fenómeno de la fragmentación de la esfera pública digital, y la tolerancia como herramienta teórica que podría resultar útil para abordar la polarización contemporánea.[1]

La investigación original partió de algunas preguntas sobre la esfera pública digital y su relación con la noción habermasiana de esfera pública, que agrupé en tres tipos.

Un primer tipo de preguntas apunta a si se ha desarrollado «un equivalente online, o sustituto online, de la aparentemente deficiente «vieja» esfera pública» (Schäfer, 2015: 322; Dean, 2003; Dahlgren, 2005). Si tenemos como objetivo una mejor comprensión de la esfera pública digital ¿qué puede aportar la noción de esfera pública acuñada por Habermas (1997) y posteriormente corregida y aumentada (Habermas, 2000; Fraser, 1990; Cohen y Arato 1992; Benhabib, 2018; Kellner, 2000)?

Un segundo tipo de interrogantes se centra en qué signo tiene una nueva transformación estructural digital de la esfera pública (Dahlberg, 1998; Feenberg y Barney, 2004; Habermas 2022). En una formulación menos técnica: Internet y las redes sociales, ¿son buenos o malos para la democracia? (Sunstein, 2018; Ilves, 2018). ¿Cuáles son sus riesgos y potenciales democráticos específicos?

De ahí deriva un tercer grupo de cuestiones: ¿Cómo lidiar con problemas como las noticias falsas y los discursos de odio? ¿Deberían regularse las redes sociales? ¿Quiénes deberían definir esa regulación: las propias plataformas, el Estado, los organismos internacionales, la sociedad civil, alguna combinación de los antedichos? (Celeste, 2018; Gill, Redeker y Gasser 2015; Iglesias Keller, 2018; Hawtin, 2011).

El artículo comienza por el problema de la fragmentación, que ha sido señalado como el principal problema de la esfera pública digital qua esfera pública por Habermas, y en relación a este problema abordo la cuestión de si internet y las redes sociales son «buenas o malas» para la democracia (Sunstein, 2018; Ilves, 2018). En segundo término, presento la conceptualización de tolerancia que ha trabajado Rainer Forst, que reformulo teniendo en cuenta la polarización contemporánea. En tercer término articulo lo anterior retomando el tema de la democracia, realizando un análisis crítico del último libro de Habermas (2022), donde analiza la esfera pública digital y postula que estamos asistiendo a una nueva transformación estructural de la esfera pública (Habermas, 1997, 2022), que podemos asociar con la fragmentación mencionada. El artículo cierra con algunas observaciones finales.

1 La referencia del título es a Habermas, 1953 ("Pensar con Heidegger, contra Heidegger").

1. Internet y las redes sociales: ¿son positivas o negativas para la democracia?

Aunque Habermas asigna un lugar muy destacado a la comunicación digital,[2] y señala algunos aspectos positivos,[3] enfatiza un aspecto negativo: su fragmentación (Carriquiry, 2022a). Ésta es crucial porque afecta al núcleo mismo de la esfera pública. En efecto, la fragmentación va en sentido contrario del proceso histórico de surgimiento de la esfera pública, y de la conceptualización que Habermas (1974, 1997) realizó reconstructivamente a partir del mismo.

Para el filósofo alemán, la esfera pública está constituida por ciudadanos reunidos formando un público: en ésta son centrales la selección de temas de interés común, su agrupamiento, su entrada en la agenda pública, y su procesamiento hasta eventualmente llegar a una opinión pública considerada, es decir reflexiva. La web muestra fallos justamente en este proceso que podríamos llamar centrípeto, de articulación y tratamiento públicos de temas de interés común. Por el contrario, se dispersa en una enorme cantidad de nichos temáticos, aislados entre sí (Habermas, 2006; Habermas, 2014; Carriquiry 2022a). Aspectos de este fenómeno —como las *echo chambers, rabbit holes, filter bubbles*[4], etc.— han sido ya analizados (Pariser 2011, Sunstein 2001, 2007, 2017).

Por otra parte Habermas ha manifestado que la esfera pública digital puede cumplir un rol especialmente relevante respecto a la deliberación,[5] pero este rol varía contextualmente. En efecto, ya en 2006[6] hace un comentario puntual sobre Internet que resulta significativo. Parte de analizar la creencia en que Internet, en comparación con los medios masivos, reintroduciría elementos deliberativos en la comunicación, basados fundamentalmente en cierta superación del carácter impersonal y asimétrico de los medios masivos. En este sentido, formula un reconocimiento relevante: «Internet ciertamente ha reactivado las bases de un público igualitario de escritores y lectores» (Habermas 2006: 423, nota 3). Esta afirmación podría leerse en clave optimista, alineada con la hipótesis de que la esfera pública digital habilita a que renazca de sus cenizas la esfera pública clásica, aquel público igualitario de escritores y lectores del siglo XVIII. Sin embargo, Habermas agrega enseguida una ponderación que aparentemente va en sentido contrario: «la comunicación mediada por computadoras en la web puede reclamar inequívocos méritos democráticos solo para un contexto especial: puede socavar la censura de los regímenes autoritarios que intentan controlar y reprimir la opinión pública» (Habermas, 2006: 423, nota 3). En estos regímenes autoritarios, la esfera pública digital puede cumplir un rol central.

---

2 Ha dicho que es una de las tres grandes innovaciones en la comunicación de toda la historia —junto con la escritura y la imprenta (Habermas, 2014).

3 A saber: apertura de la posibilidad de ser autores, acceso de cada vez más personas a cada vez más información, y posibilidad de generación de lo que ha conceptualizado como opinión pública considerada.

4 Cámaras de eco, madrigueras o agujeros de conejo, y filtro burbuja o burbujas de filtro.

5 Para un tratamiento agudo y ya clásico sobre el alcance preciso de la deliberación, incluyendo las modulaciones entre «deliberar» o «debatir» y negociar» véase Elster (2000).

6 En el texto de una conferencia que dictó en 2006 (Habermas, 2006), más precisamente en una nota al pie. Cabe mencionar que también fue conferencista invitado en el mismo evento Manuel Castells (Castells, 2008).

En cambio, para el contexto de regímenes liberales, el rasgo que Habermas destaca en dicho trabajo es justamente la fragmentación. Las grandes audiencias masivas pero políticamente enfocadas se ven fragmentadas en un gran número de «públicos» focalizados en temáticas específicas y aislados entre sí (Habermas, 2006: 423, nota 3). En este contexto Habermas relativiza el alcance de los debates en línea, afirmando que solo promueven la comunicación política cuando cristalizan alrededor de los puntos focales de la prensa independiente de calidad (Habermas, 2006: 423, nota 3).

En trabajos posteriores, Habermas ha mantenido estas dos ideas fuerza sobre Internet: la fragmentación por un lado, y su potencial en regímenes totalitarios por otro. En una entrevista de 2016 enfatiza la fuerza centrífuga de la web, que libera una ola anárquica de circuitos de comunicación altamente fragmentados, y al mismo tiempo reconoce que la espontaneidad y la falta de límites puede tener efectos subversivos en contextos autoritarios. Y subraya: «la web en sí misma no produce esferas públicas» (Habermas 2016b). A primera vista esta frase puede impresionar como lapidaria, pero un análisis más detenido puede llevar a preguntarnos si, incluso si aceptamos que Internet «en sí mismo» no produce esferas públicas —como, mutatis mutandis, tampoco las *coffee houses*[7] del siglo XVIII «en sí mismas» produjeron esferas públicas—, es posible pensar qué otros elementos serían necesarios para desplegar esferas públicas con el concurso de la web.

En esta línea, la cuestión de qué rol puede cumplir Internet en las democracias contemporáneas, en términos habermasianos se formularía como qué rol puede cumplir la esfera pública digital en el marco de un modelo deliberativo de democracia.[8] En este sentido creo que hay que tener en cuenta dos elementos. En primer lugar, que lo que dice Habermas (2006) es que Internet solo puede atribuirse «inequívocos méritos democráticos» en contextos autoritarios. Es decir que, en principio, la teoría habermasiana no excluye que Internet pueda tener algo así como «ambivalentes méritos democráticos» en otros contextos, incluyendo las democracias liberales contemporáneas.

En segundo lugar, hay que tener en cuenta que las primeras afirmaciones citadas son de 2006, y que desde entonces no solo la esfera pública digital ha cambiado aceleradamente —en particular, las «salas de chat» o «grupos de noticias» que analizaba Habermas han sido desplazadas por las redes sociales—, sino que también el filósofo alemán ha evolucionado en su reflexión sobre el tema. Ha seguido advirtiendo contra los peligros de la fragmentación —y la creciente polarización parece haberle dado la razón—, pero también ha mostrado cierta apertura a rasgos potencialmente positivos.

Identifico esta apertura en dos aspectos. Por una parte en su noción más general de procesos de aprendizaje colectivo, que en este tema se manifiesta al respecto de ser «autores en potencia» y del rol de la prensa en un aprendizaje «autopaternalista» de los lectores (Habermas 2007)[9]; retomaré este punto más adelante. Por otra parte, en el reconocimiento de la limitación de su experiencia personal: en una entrevista más reciente declaró que «soy demasiado viejo para juzgar el impulso cultural que originarán los nuevos medios» (Habermas, 2018). Más que a una limitación meramente generacional se refiere a la particularidad

7 Ni los salones ni las *Tischgesellschaften*.

8 Desplegado por Habermas (2000) en contraposición a los modelos republicano y liberal.

9 Un tratamiento de este punto en Carriquiry, 2022a.

de su experiencia en el área: en otra entrevista había reconocido que «dado que uso Internet solo para objetivos específicos y no muy intensivamente, no tengo experiencia de redes sociales como Facebook y no puedo hablar del efecto generador de solidaridad de la comunicación electrónica, si es que existe» (Habermas, 2016b). Creo que este reconocimiento de sus propias limitaciones habilita un espacio potencial donde pensar con Habermas más allá de Habermas.

En este marco inscribo la respuesta de Habermas a la pregunta: «¿Internet es beneficioso o perjudicial para la democracia?». Lo primero que responde es: «Ni una cosa ni la otra» (Habermas, 2014). Luego argumenta en líneas convergentes con lo analizado hasta aquí, subrayando la enorme importancia histórica de Internet en relación al acceso al conocimiento y a volvernos autores, pero a la vez alertando sobre su tendencia a la fragmentación, a lo que agrega el rol que puede cumplir el periodismo para contrarrestar dicha tendencia (Habermas, 2014; Carriquiry, 2022a).

En este sentido identifico algunas similitudes con el trabajo de Cass Sunstein (2018) «¿Las redes sociales son buenas o malas para la democracia?», en el que sintetiza algunas ideas presentes en trabajos anteriores (Sunstein 2001, 2007 y 2017). Para empezar, la pregunta es bastante similar a la que le formulaban a Habermas, aunque la planteada por Sunstein es menos general, enfocándose en las redes sociales. Además, ambas preguntas tienen una formulación dicotómica, que parece expresar cierta ambivalencia con respecto a Internet y las redes sociales, y cierta urgencia por eliminar esa ambivalencia.

La respuesta de Sunstein, cuatro años posterior a la de Habermas, también es semejante en algunos puntos. En primer lugar porque, pese al carácter dicotómico de las preguntas, ambos autores se niegan a dar una respuesta unilateral, y en cambio marcan aspectos positivos y negativos. En segundo lugar porque ambos coinciden al destacar, entre los rasgos positivos, el mayor acceso a la información. Sin embargo, difieren al identificar el aspecto negativo central: Habermas señala la «fragmentación» y Sunstein la «polarización». En lo que sigue profundizaré en aspectos de ambos problemas. Respecto a la polarización planteo una conceptualización de la noción de tolerancia -basada en Forst aunque reformulada- como herramienta teórica para abordar el fenómeno, incluyendo los discursos de odio como una expresión del mismo, y articulo dicha conceptualización con la "polarización democrática" que ha propuesto Habermas (2016). Respecto a la fragmentación realizo un análisis crítico del último libro de Habermas (2022), donde presenta una visión más pesimista centrada justamente en este problema.

En cualquier caso, la cuestión más acuciante parece ser cómo contrarrestar o mitigar estos aspectos negativos y fortalecer los positivos. Para contribuir a una posible respuesta a esa cuestión, además de la conceptualización de tolerancia, creo útil aplicar a la esfera pública digital otra herramienta conceptual que funge como complemento: la noción de procesos de aprendizaje colectivo. En esta línea recupero algunos conceptos vertidos por Habermas (2018) sobre un posible proceso de aprendizaje social con respecto a las redes sociales. En dicha entrevista, Habermas reafirma la idea ya mencionada de que Internet convirtió a los lectores en autores, y la amplía en un sentido ligeramente más optimista al mencionar la posibilidad de un proceso de aprendizaje análogo al que nos convirtió en lectores a partir de la invención de la imprenta. El entrevistador le pregunta «¿No cree que Internet, más allá de sus indiscutibles ventajas, ha forjado una especie de nuevo analfabetismo?» (Habermas,

2018). Habermas comienza mencionando justamente las problemáticas de la polarización y la fragmentación: «Usted se refiere a las controversias agresivas, las burbujas y los bulos de Donald Trump en sus tuits» (Habermas, 2018). Pero enseguida enmarca este fenómeno en una mirada histórica más larga, retomando el paralelismo entre las innovaciones constituidas por la imprenta y la web (Habermas, 2014). Subraya entonces que, lejos de ser un proceso instantáneo, pasaron siglos desde la invención del libro impreso, que convirtió a toda la población en lectores en potencia, hasta que efectivamente todas las personas aprendieron a leer —o al menos una gran mayoría. Internet, por su parte, que «nos convierte a todos en autores en potencia, no tiene más que un par de décadas de edad. Es posible que con el tiempo aprendamos a manejar las redes sociales de manera civilizada» (Habermas, 2018).

2. La tolerancia como herramienta conceptual para abordar la polarización

En este marco propongo inscribir una conceptualización de la tolerancia basada en la que hace Rainer Forst (2013, 2014, 2017), figura contemporánea de la Teoría Crítica, pasible de ser articulada con la teoría habermasiana de la democracia deliberativa.

El término «tolerancia»[10] para Forst refiere a «la aceptación condicional o no interferencia con creencias, acciones o prácticas que uno considera que son incorrectas, pero aún así "tolerables", de manera que no deben ser prohibidas o restringidas» (Forst, 2017). Es importante destacar aquí un rasgo muy relevante, a saber, que la tolerancia aplica a un subconjunto de las creencias o prácticas con las que no estamos de acuerdo: aquellas que consideramos incorrectas, pero no «intolerables», es decir, aquellas que no prohibiríamos. Es importante también subrayar que el alcance de lo tolerable dependerá también del contexto específico de que se trate, como veremos (Forst, 2017).

Para Forst se puede identificar tres componentes en el concepto de tolerancia. El primero es el que adelantábamos recién: el componente de objeción, que es crucial. Para poder hablar de tolerancia es necesario que las creencias o prácticas toleradas se consideren objetables «y, en un sentido importante, incorrectas o malas» (Forst, 2017). De lo contrario hablaríamos, o bien de adhesión (en relación a creencias o prácticas con las que se está de acuerdo), o bien de indiferencia (si las creencias o prácticas no despiertan ni adhesión ni objeción).

El segundo componente es el de aceptación, que equilibra a la objeción. Este ingrediente de aceptación, sin anular el componente de objeción, da ciertas razones positivas que superan a las negativas. A la luz de las primeras «sería incorrecto no tolerar lo incorrecto» (Forst, 2017).[11] Es decir que las creencias o prácticas toleradas se consideran incorrectas, pero no intolerablemente incorrectas. (A modo de ejemplo: esto podría darse en el caso de una persona religiosa que tenga la creencia de que las prácticas homosexuales son incorrectas, pero las tolera porque esgrime como razón positiva la no-discriminación de las personas homosexuales.)

El tercer componente es el de rechazo, que refiere a lo intolerable, es decir, a los límites de la tolerancia, que deben ser especificados (Forst, 2017). Esta frontera donde termina

10 Como es sabido, la noción de tolerancia tiene una rica historia, que no cabe desplegar en este artículo: un panorama enciclopédico en el propio Forst (2017).

11 Por mencionar una conocida paradoja de la tolerancia (que Forst analiza junto a otras dos).

la tolerancia y comienza lo intolerable se determina por el punto donde hay razones para el rechazo que son más fuertes que las razones para la aceptación. Como aclara en otro trabajo, las primeras deben ser «particularmente sólidas si conllevan consecuencias legales en el sentido de restringir la libertad» (Forst, 2014: 67). (Un ejemplo ilustrativo en este sentido podrían constituirlo las fuertes restricciones en Alemania respecto al negacionismo del Holocausto. Podríamos decir que, en este caso, la sociedad alemana ha decidido que las razones para rechazar el negacionismo son "particularmente sólidas". En cualquier caso, el filósofo alemán explicita que queda abierta la cuestión de los medios apropiados para una posible intervención (Forst, 2017).

Esta última precisión podría aplicarse con provecho al caso de la esfera pública digital, donde la cuestión de los medios apropiados para una intervención con respecto a, por ejemplo, los discursos de odio, es problemática y abarca desde la autorregulación de las propias empresas —en base a denuncias de los usuarios y/o en base a «Terms of use» que los usuarios han suscrito— hasta la regulación de los Estados o de organismos multilaterales, pasando por iniciativas de la sociedad civil.

Como aclara en otro trabajo, estamos lidiando con tres clases de razones: para objetar una creencia o práctica, para aceptar una creencia o práctica que objetamos, y para rechazar una creencia o práctica que objetamos. El ejercicio de la tolerancia siempre implica «la tarea de conectar estas razones de la manera correcta» (Forst, 2014: 67); de ahí su complejidad, pero también sus potencialidades.

Creo que esta conceptualización, y en particular la coexistencia entre objeción y aceptación en una tensión fructífera, es particularmente consistente con algo que Habermas señala como el derecho a seguir siendo extraños[12]. Habermas sostiene que en una sociedad secularizada que ha aprendido a lidiar con su complejidad consciente y deliberadamente, el dominio comunicativo de los conflictos constituye la única fuente de solidaridad entre extraños que renuncian a la violencia y, en la regulación cooperativa de su vida en común, también se conceden mutuamente el derecho a seguir siendo extraños (Habermas, 2000: 385-6).[13] Habermas engarza esta noción en el conjunto de su teoría de la sociedad, vinculándola con esta posibilidad de tramitación comunicativa de los desacuerdos y conflictos propios de las sociedades complejas.

12 Aunque Habermas aquí se refiere a "extraños" en un sentido general, podría aplicarse también a los "extranjeros" en particular; para este tema véase el trabajo de Juan Carlos Velasco sobre migrantes y refugiados (Velasco, 2019).

13 Este énfasis, que va a reaparecer en su último libro (Habermas, 2022), converge llamativamente con algo que el propio Habermas había señalado cuando, para resumir la idea de reconciliación de Adorno, cita el siguiente pasaje de su antiguo mentor: «El estado reconciliado no se anexionaría lo extraño con imperialismo filosófico, sino que pondría su alegría en que en la proximidad que le concede siguiera siendo lo lejano y diverso, más allá tanto de lo heterogéneo como de lo propio» (Adorno, 1973, VI: 192, citado en Habermas 1999: 498). Creo que hay un eco de este conceder seguir siendo «lo lejano y diverso» de Adorno, en el énfasis que hace Habermas en que estos extraños se conceden mutuamente el derecho a seguir siendo extraños.

2.1. Tolerancia en tres tipos de sociedades

Por otra parte Forst delinea tres ámbitos de creencias y prácticas. A saber: 1) las creencias y prácticas con las que uno está de acuerdo, 2) las creencias y prácticas con las que uno está en desacuerdo pero puede tolerar, y 3) las creencias y prácticas intolerables, es decir con las que uno está en desacuerdo y no puede tolerar.

Represento gráficamente estos tres ámbitos como un rectángulo dividido en tres partes (ver figura 1) donde la parte central representa al ámbito (2) y las otras dos partes representan respectivamente a (1) y (3).

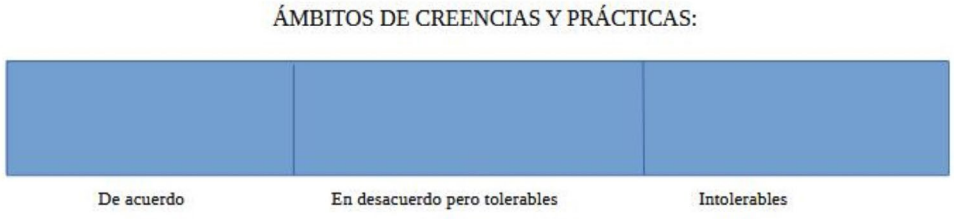

Figura 1. Creencias y prácticas según tolerancia (Fuente: Elaboración propia).

Propongo entonces representar distintos tipos de sociedades según cómo podrían distribuirse estos tres ámbitos. En primer lugar, se podría identificar una sociedad de tipo tradicional en la que el ámbito (1) estuviera muy desarrollado, es decir en la que sus ciudadanos están de acuerdo en una gran proporción de creencias y prácticas (ver figura 2). (Un ejemplo ilustrativo de este tipo podría constituirlo, por caso, una sociedad teocrática y sin disidencia interna).

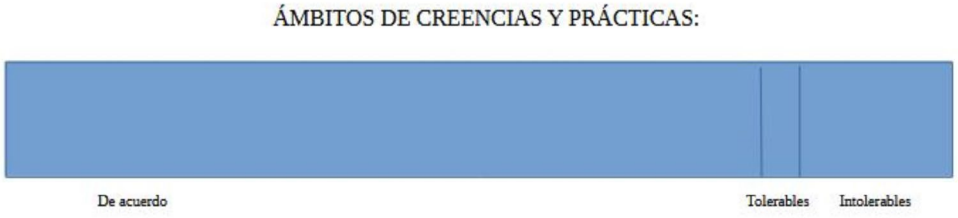

Figura 2. Sociedad de tipo tradicional (Fuente: Elaboración propia).

En segundo lugar, se podría representar una sociedad de tipo complejo y pluralista en que el ámbito (2) estuviera muy desarrollado, es decir en la que haya una gran proporción de creencias y prácticas sobre las cuales sus ciudadanos están en desacuerdo, pero toleran. Correlativamente, los ciudadanos estarían de acuerdo en una proporción relativamente pequeña de creencias y prácticas (reflejado gráficamente en un espacio acotado del ámbito

(1)), y también sería pequeño el ámbito de creencias y prácticas que son consideradas intolerables (ver figura 3). (Como caso ilustrativo podríamos pensar en los EEUU de la última década del siglo xx, más precisamente en la autocomprensión que tenía esa sociedad de sí misma como "crisol" de diferentes cosmovisiones —por supuesto, es posible desplegar una visión crítica sobre dicha autocomprensión; en este punto quedémonos con el ejemplo a modo de mera ilustración).

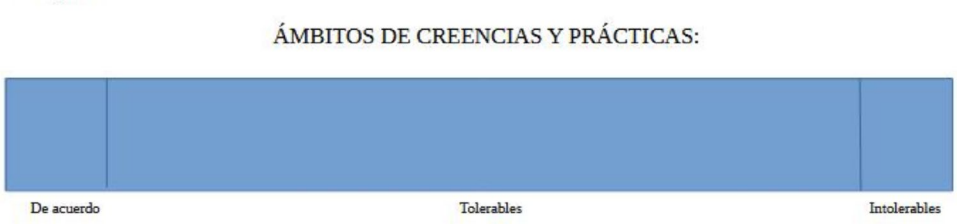

Figura 3. Sociedad de tipo «tolerante» (Fuente: Elaboración propia).

Finalmente podría representarse una sociedad muy polarizada en que el ámbito (2) fuera muy reducido, es decir que habría muy pocas creencias y prácticas con las que sus integrantes están en desacuerdo, pero toleran. Esto podría redundar en espacios más extendidos para el ámbito (1) y el (3). Si además quisiéramos representar una sociedad compleja donde la proporción de creencias y prácticas en las que sus ciudadanos están de acuerdo es relativamente acotada, graficaríamos el ámbito (1) como menor al ámbito (3) (ver figura 4). (Un ejemplo podría constituirlo una sociedad del tipo de la argentina en los últimos años, donde la noción de "la grieta" —entre visiones políticas que en cierto sentido se consideran inconmensurables— podría interpretarse como sintomática de este tipo de polarización).

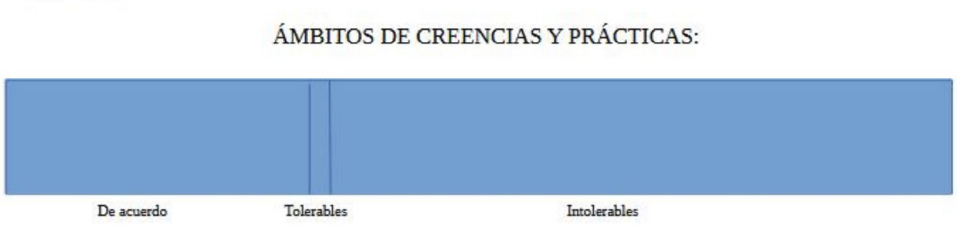

Figura 4. Sociedad de tipo polarizada (Fuente: Elaboración propia).

2.2. Desacuerdo, discursos de odio y polarización democrática

Por supuesto, es posible postular otras variantes y combinaciones, y por otra parte la representación gráfica tiene límites heurísticos que no permiten capturar otros aspectos. Lo que quiero enfatizar con las variantes que presenté es la relación entre el fenómeno

de la polarización y la amplitud relativa del ámbito de creencias y prácticas toleradas. En efecto, conceptualizar el ámbito de lo tolerable como aquello con lo que estamos en desacuerdo pero sin embargo podemos tolerar da cuenta del desacuerdo que efectivamente existe en sociedades complejas, y al mismo tiempo permite imaginar el posible desarrollo de la tolerancia hacia un conjunto más amplio de esas creencias y prácticas sobre las que existe desacuerdo.

Un proceso de polarización creciente tendría el mismo punto de partida pero se desplegaría en sentido opuesto a la tolerancia. Es decir que tiene como base el desacuerdo sobre una gran proporción de las creencias y prácticas, pero gestiona ese desacuerdo de modo tal que las creencias y prácticas con las que se está en desacuerdo son absorbidas por el ámbito (3) es decir, pasan a ser intolerables.

En ese sentido, una de las virtudes de la conceptualización de tolerancia de Forst, que comparto, es que enfatiza el desacuerdo como presupuesto para la tolerancia. Esto es especialmente relevante dado el desacuerdo como rasgo característico de las sociedades contemporáneas. En esta línea, el desacuerdo no se resuelve ni se disuelve: se reconoce su existencia, y al mismo tiempo se abren las dos posibilidades representadas por su elaboración como tolerable o intolerable.

En síntesis: si partimos de la base de un tipo de sociedad compleja en la que existe desacuerdo sobre una proporción importante de creencias y prácticas, se pueden esquematizar dos alternativas posibles para tramitar ese desacuerdo. O bien es absorbido por el ámbito (2), es decir que las creencias y prácticas con las que uno está en desacuerdo resultan tolerables, o bien es absorbido por el ámbito (3), es decir que resultan intolerables. El resultado de estas dos alternativas puede representarse en las figuras 3 y 4 respectivamente.

La clave para que una creencia o práctica eventualmente pueda pasar de ser intolerable a ser tolerable es el peso específico del segundo componente que plantea Forst, es decir el de las razones positivas que equilibran el componente de objeción. (En el ejemplo citado ut supra, de una persona religiosa que tiene la creencia de que las prácticas homosexuales son incorrectas, tendría dos alternativas. Si pesaran más sus razones para el rechazo de dichas prácticas, éstas resultarían intolerables y la persona podría apoyar mecanismos legales que restringieran la libertad de las personas homosexuales. En cambio, esas prácticas pasarían a ser tolerables si pesaran más las razones positivas, por ejemplo la importancia de no discriminar a las personas homosexuales).

En este marco podría pensarse una respuesta posible a la tensión entre la libertad de expresión y los discursos de odio. Como adelanté, los discursos de odio son uno de los fenómenos problemáticos de la esfera pública digital —que la desbordan, ya que también se extienden fuera del ámbito digital. Una de las dificultades conceptuales radica en la tensión problemática entre lidiar con los discursos de odio y preservar la libertad de expresión, que muchas veces se expresa en si se debería o no censurar los discursos de odio, lo que puede remitir al problema clásico y más general de si se debería tolerar al intolerante.

En la línea que venimos planteando, una respuesta podría ir en la línea de «arrinconar» a estos últimos como intolerables. Es decir, identificar los discursos de odio con lo intolerable. Pero ese movimiento requiere a la vez que se pueda tolerar —implicando escuchar y debatir— el muy amplio espectro de lo que no es discurso de odio. Esto implicaría entonces

un uso preciso y restringido de la categorización, ya que, si se extiende indiscriminadamente la aplicación de esa etiqueta, se opera en el sentido de una mayor polarización —es decir, hacia un escenario del tipo esquematizado en la figura 4.

Una estrategia en esta línea podría implicar entonces ampliar en todo lo que sea posible el alcance del ámbito (2), es decir las creencias y prácticas que toleramos, y complementariamente hacer menos énfasis en el límite entre los ámbitos (1) y (2), es decir entre las creencias y prácticas con las que estamos de acuerdo y aquellas con las que estamos en desacuerdo pero toleramos. Una vez conseguido esto, se podría enfatizar muy fuertemente el límite entre los ámbitos (2) y (3), es decir el límite donde termina la tolerancia.

De este modo podría pensarse que una mayor tolerancia permitiría marcar con más fuerza lo realmente intolerable. Dado que, en un escenario altamente polarizado donde una gran proporción de creencias y prácticas resultan intolerables, cuando aparece una creencia o práctica que resulta aún más intolerable que las demás, es muy difícil señalarlo —metafóricamente: ya no se puede gritar más alto. Esto resulta especialmente relevante para la esfera pública digital, donde la tensión entre la libertad de expresión y los discursos de odio es agudamente problemática. Esto implica distintos factores que no podemos analizar en detalle aquí pero que cristalizan por ejemplo en las críticas a los mecanismos de moderación de contenidos en redes sociales.[14]

En un sentido más general, esta propuesta de abordaje de la polarización se puede articular con lo que Habermas ha llamado una «polarización democrática» (Habermas, 2016; Carriquiry, 2019). Ésta se diferencia de lo que en general se denomina como polarización (que en lo que sigue propongo denominar «polarización extrema»), y podríamos explicarla sintéticamente en dos movimientos. A saber: como primer movimiento se denuncia a las tendencias antidemocráticas como tales —incluyendo a las derechas populistas—, para luego tender a una «polarización democrática» interna por decirlo así, entre propuestas de derecha democrática y de izquierda, en la cual el papel de ésta última debería ser propositivo. En cualquier caso, creo que puede resultar esclarecedora la diferencia entre una «polarización extrema» y esta «polarización democrática». La noción de tolerancia, lejos de excluir este último tipo de polarización, de algún modo la implica. En efecto, lo tolerable no es lo que nos resulta indiferente, sino aquello con lo que estamos en desacuerdo —y que sin embargo podemos tolerar. Una polarización extrema involucra un escenario en el cual el desacuerdo se aborda de modo tal que resulta en una baja proporción de creencias y prácticas toleradas, y una gran proporción de creencias y prácticas intolerables. Una «polarización democrática» cumpliría con tres condiciones compatibles con la conceptualización de tolerancia que presentamos: lejos de negar el desacuerdo, lo enfatiza; no convierte una gran proporción de creencias y prácticas en intolerables; y sí señala como intolerables a aquellas posiciones que, como el populismo de derechas, constituyen, en palabras de Habermas (2016), «un caldo de cultivo para un nuevo fascismo».

14 Un tratamiento de esta cuestión en Frost-Arnold, 2023, una ilustración en Block y Riesewieck, 2018.

3. Una nueva transformación de la esfera pública

En su libro más reciente, Habermas (2022) vuelve sobre el tema de la esfera pública y la teoría de la democracia deliberativa, y en ese marco aborda la esfera pública digital.[15] Recoge así algunos puntos que hasta entonces habían estado dispersos y suma otros nuevos. En lo que sigue reviso los relevantes a los efectos de este artículo, para luego cerrar con una valoración crítica que apunta a seguir pensando estos problemas abiertos.

Entre los puntos que Habermas reafirma en su nuevo libro, y que habíamos mencionado ut supra, aparecen: la fragmentación de la esfera pública asociada a las plataformas digitales; la calificación del surgimiento de internet como una de las tres grandes revoluciones de la comunicación en la historia entera de la humanidad; la idea de que todos nos hemos vuelto autores, y el énfasis en el tiempo que ese proceso de aprendizaje implicará; el rol neurálgico de la prensa en las sociedades complejas.[16]

Entre los aspectos relativamente nuevos que aparecen, algunos son cuestiones de énfasis. Respecto a la prensa, enfatiza aún más su rol crucial, señalando que la prensa genera, a partir de diferentes interpretaciones del mundo que compiten entre sí, un núcleo de interpretación intersubjetivamente compartido. Los medios de comunicación, «con su flujo de información e interpretaciones actualizadas diariamente, confirman, corrigen y complementan constantemente la borrosa imagen cotidiana de *un mundo que se supone objetivo* y que más o menos *todos los contemporáneos* suponen también aceptado como "normal" o válido por todos los demás» (Habermas, 2022: 55).[17]

También es nuevo, por supuesto, lo que aparece desde el título del libro:[18] Habermas sostiene ahora claramente que ha habido una nueva transformación estructural de la esfera pública, en la que la revolución digital ha tenido un impacto significativo.

En este marco, plantea un diagnóstico sombrío. Por una parte, entre quienes utilizan exclusivamente las redes sociales parece estarse desplegando un tipo de comunicación que denomina como «semi-pública», que pone en riesgo la formación deliberativa de opinión y de voluntad políticas: «entre los usuarios exclusivos de las redes sociales parece estar ganando aceptación una forma de comunicación semipública, fragmentada y circular, que está deformando su *percepción de la esfera pública política* como tal. Si esta suposición es correcta, se está poniendo en peligro un importante requisito subjetivo para el modo más o menos deliberativo de formación de la opinión y la voluntad entre una proporción creciente de ciudadanos» (Habermas, 2022: 11-12).

En este aspecto radica el vínculo entre la fragmentación de la esfera pública y el riesgo para la democracia. En efecto, para Habermas la diferencia de la democracia moderna respecto a la antigua radica en la igualdad de derechos subjetivos que otorga a sus ciudadanos

15 El libro tiene tres capítulos: el primero se basa en un artículo que Habermas había publicado en 2021, el segundo consiste en una entrevista a Habermas que había sido publicada en 2018, y el tercero se basa en un texto que había sido publicado en 2022, ampliado con consideraciones adicionales. En el momento en que escribo este artículo, el libro como tal aún no ha sido traducido al inglés ni al español.

16 Un tratamiento de este último punto en Carriquiry, 2022a.

17 En esta cita y todas las de Habermas, 2022, la traducción es mía.

18 «Una nueva transformación estructural de la esfera pública y la política deliberativa».

y en su carácter representativo, que hace que la voluntad política de los ciudadanos solo pueda ejercerse indirectamente, es decir, a través de elecciones generales.

A nuestros efectos, lo relevante es justamente que este acto de voluntad ejercido conjuntamente solo puede cumplirse en una esfera pública inclusiva. Es decir que solo si los actos electorales surgen de la participación de los ciudadanos en una comunicación común, esas decisiones serán «hechas individualmente e independientemente por cada cual como resultado de un proceso de toma de decisión común» (Habermas, 2022: 89). Es por esto que Habermas sostiene que el carácter público —es decir, no semipúblico— de la comunicación, forma el vínculo necesario entre por un lado la autonomía política de cada individuo y por otro lado la voluntad política común de toda la ciudadanía. En síntesis:

Solo como participante en el proceso de formación de la opinión pública puede el ciudadano individual, en su formación individual de opinión y en su toma de decisiones, equilibrar la tensión que existe entre los intereses individuales del ciudadano de la sociedad y el bien común del ciudadano del Estado (Habermas, 2022: 89).

Dicho de otro modo: encontrar el equilibrio entre el ejercicio interesado de las libertades subjetivas y la orientación hacia el bien común, que es funcionalmente necesaria, debe ser resuelto por los propios ciudadanos, y solo puede resolverse en el proceso de formación conjunta de la opinión y la voluntad en la esfera pública política (Habermas, 2022: 95). En este sentido converge con interpretaciones como la de Ipar cuando afirma que Habermas intenta «abrir el horizonte normativo desde el cual confrontar contra la apropiación instrumental e individualista de las instituciones de la libertad y la democracia modernas» (Ipar, 2021: 225).

Para Habermas, la esfera pública democrática está entonces amenazada por el efecto acumulativo de acontecimientos interdependientes. Ante esta situación, señala una doble carencia: de conciencia pública de que hay un problema, y de regulación pública que podría mitigar el problema. En distintos pasajes del libro Habermas llama la atención sobre la necesidad de regulación de la infraestructura digital en general y las plataformas en particular, criticando que actualmente dicha normativa está ausente (Habermas, 2022: 108) y/o es inadecuada (Habermas, 2022: 29). La relevancia de que exista una regulación adecuada es superlativa, ya que podría contribuir al objetivo antedicho de movilizar la formación de opinión y voluntad políticas.

En síntesis, la postura de Habermas en este trabajo de 2022 es relativamente más pesimista que en trabajos anteriores que analizamos ut supra (Habermas, 2006, 2007, 2014, 2016, 2018). Este pesimismo deriva de que en los últimos años detecta no solo cambios políticos circunstanciales, como los asociados a las nuevas derechas populistas (Habermas 2016), sino un cambio *estructural* en la esfera pública, asociado a la fragmentación.

Si en su primera obra importante (Habermas, 1997) era más pesimista, y en los años noventa tuvo un giro menos pesimista, asociado al interés por los movimientos sociales (Habermas, 2000), en 2022 su visión vuelve a ser relativamente más pesimista. Si en los años sesenta postuló que la esfera pública había sufrido una transformación estructural de signo negativo, que denominó como un proceso de «refeudalización» (Habermas, 1997), en 2022 sostiene que ha habido una nueva transformación estructural de la esfera pública, también de signo negativo, que propongo denominar como un proceso de «fragmentación».

De todos modos Habermas, a diferencia de la primera generación de la Teoría Crítica —quizás justamente en contraposición a ella— siempre mantiene abierta la posibilidad de que

haya una luz al final del túnel. Algo de esto puede encontrarse en algunos puntos de su último libro: por ejemplo cuando afirma que la desintegración de la esfera pública política podría ser «temporal» (Habermas, 2022: 64), o cuando señala que la esfera pública digital ha sido también utilizada por las «valientes mujeres bielorrusas» (Habermas, 2022: 46) en sus protestas. Y, más en general, en una actitud propositiva, que como vimos centra en la regulación política.

4. Observaciones finales

Por mi parte, en líneas generales coincido en el diagnóstico de Habermas, particularmente con los agudos dardos críticos hacia los gigantes tecnológicos:[19] «la lava del potencial simultáneamente antiautoritario e igualitario, que aún podía sentirse en el espíritu emprendedor californiano de los primeros años, pronto se solidificó en Silicon Valley en la mueca libertaria de las corporaciones digitales que dominan el mundo» (Habermas, 2022: 46).

Sin embargo, hay que señalar que el planteo de Habermas presenta algún flanco débil, por ejemplo en relación con fenómenos que aborda solo indirectamente. Es el caso de los influencers y lo que Habermas (2022) denomina como «autopresentación narcisista» (Habermas, 2022: 58), que he propuesto interpretar, respectivamente, a la luz de la noción de influencia y la privacidad orientada a una audiencia (Carriquiry, 2022b).

«Si se distingue. claramente la «individualización» de la «singularización», es decir, el carácter distintivo adquirido a lo largo de la vida de una persona, de la visibilidad pública y la ganancia de distinción que puede obtener mediante apariciones espontáneas en línea, por ejemplo, la «promesa de singularización» puede ser el término adecuado para los influencers que buscan la aprobación de los seguidores para su propio programa y su propia reputación» (Habermas, 2022: 58).

En este punto cabría preguntarse si la lectura de Habermas está dando cuenta cabalmente del hecho de que actualmente, para las nuevas generaciones, su actividad en las redes sociales es parte inextricable de su proceso de individualización, ese carácter distintivo adquirido a lo largo de la vida. Aunque para ser justos hay que aclarar que en este punto Habermas está enfocándose en los influencers, también hay que hacer notar que existen diferentes tipos de influencers, y Habermas parece estar pensando solo en el tipo que monetiza esa influencia.

En este y otros puntos, mi lectura es ligeramente menos pesimista que la de Habermas. En efecto, creo que es posible construir una interpretación que haga pie en que la idea emancipatoria de igualdad asociada a internet funge como «aguijón interno»: como idea, permanece operando en la realidad de la que es parte. En ese sentido, creo que si bien es indudable la relevancia de las *eco chambers, rabbit holes, filter bubbles*, etc., la esfera pública digital no se reduce a ellas. Y sobre todo creo que hay mecanismos «autopaternalistas» que nos han sido útiles en el pasado y que pueden fungir como auxiliares para atravesar este proceso de aprendizaje hacia ser autores. Uno de esos mecanismos es, como he analizado en este artículo, la tolerancia.[20]

19 Las grandes empresas tecnológicas —Google, Apple, Meta, Amazon, Microsoft—, cuyo crecimiento las ha llevado a ocupar los puestos más altos en los rankings globales empresariales, llegando a superar por ejemplo a las del área energética.

20 Otro de esos mecanismos, como he planteado en otro trabajo (Carriquiry, 2022a) es fortalecer la prensa independiente. Este fortalecimiento puede tomar varias formas. Una es las diferentes formas de apoyos o subsidios estatales (Carriquiry, 2022a). Otra posibilidad es fomentar nuevas iniciativas del tipo de The

En efecto, creo que la conceptualización de tolerancia que presenté —con Forst y más allá de Forst— puede articularse adecuadamente con un modelo de democracia deliberativa, y específicamente con una noción de esfera pública como espacio que habilita el dar y pedir razones para las creencias y prácticas, señalar buenas razones respecto a determinada postura y también denunciar las malas razones, proveer informaciones pertinentes y aportes relevantes en general (Habermas, 2000; Cohen y Arato, 1992). Esta articulación es especialmente adecuada dado el énfasis de Forst en el sopesar las tres clases de razones —para la objeción, la aceptación o el rechazo—. Uno de los mecanismos por el cual, en cada caso de desacuerdo sobre una creencia o práctica determinada, ésta puede pasar de ser intolerable a ser tolerable, coincide con el procedimiento enfatizado en un modelo deliberativo de democracia: el debate público racional que puede tener lugar en la esfera pública dando como resultado una opinión pública reflexiva.

En este sentido, la esfera pública digital presenta fortalezas y debilidades significativas. Entre las fortalezas se puede señalar, además de la mayor democratización en el acceso a la información, el hecho de que vuelve posible cierto intercambio de argumentos. Este intercambio se puede identificar en dos sentidos: en el que señala Habermas cuando afirma que ahora todos somos autores, es decir que habilita un rol más activo de los ciudadanos en la exposición de sus creencias; y también en cuanto a que algunas redes sociales habilitan espacios para el diálogo interpersonal en los «comentarios» a esas piezas autorales, relativamente más amplios a los que permitía un medio masivo de comunicación previo a la aparición de Internet (Carriquiry, 2022b). Entre las debilidades, en las que como vimos, han coincidido desde perspectivas diversas Habermas y Sunstein, entre otros, se debe enfatizar el rasgo de fragmentación ya mencionado, que puede dar lugar a una polarización más extrema.

De todos modos, en este punto del debate público radica la respuesta a la posible pregunta sobre cómo aplicar estos conceptos -amén de los ejemplos incluidos a lo largo del artículo, que intentan ilustrar posibles aplicaciones en la práctica-; más precisamente, la cuestión de cómo se podría incitar la dinámica de ampliar el ámbito de las creencias con las que estamos en desacuerdo pero consideramos tolerables[21]. En efecto, el mecanismo que propone la teoría de la esfera pública es justamente procesar las creencias y prácticas en la esfera pública; es decir que el curso de acción recomendable desde este marco es involu-

Conversation (2024), donde expertos, periodistas y editores colaboran en un modo innovador. Una tercera posibilidad es mediante regulaciones sobre las plataformas, cuyos algoritmos podrían modificarse de modo de dar más espacio a la prensa independiente: por ejemplo, que los feeds muestren más prensa de calidad. Esta es una propuesta concreta, que no implica ni censurar las fake news, ni la moderación de contenido, y en ese sentido puede sortear parte de las objeciones que se hacen en nombre de la libertad de expresión. Por otra parte, es relativamente menos costoso, y podría eventualmente conseguir apoyo de los medios de comunicación tradicionales (prensa escrita pero también grandes cadenas de televisión y radiodifusión), y quizás también de los gobiernos, en tanto da mayor estabilidad al sistema político, reduciendo el tipo de competencia frenética asociada a qué partido consigue más bots.

21 Cabe aquí una aclaración de alcance metodológico: en este trabajo me enfoco en aspectos teóricos tal como son desarrollados en, o se pueden proyectar y articular a partir de, los trabajos de Habermas y Forst. En este sentido, debo explicitar aquí un supuesto muy general del que parto: la concepción del conocimiento como una construcción colectiva, de la que este trabajo es apenas una pieza, centrada en proveer herramientas teóricas (Carriquiry 2022b). Esto implica que esta pieza debe por supuesto ratificarse o rectificarse con trabajo empírico especializado.

crarse en el debate público. En este punto no es posible formular recetas ahistóricas: dónde trazar el límite entre lo tolerable y lo intolerable sólo puede determinarlo cada sociedad, para cada caso, en cada momento histórico. (Por ejemplo, las limitaciones que ha establecido para sí Alemania respecto al negacionismo del Holocausto, no serían fácilmente aceptadas en Estados Unidos, dada su conceptualización de los alcances de la libertad de expresión; por poner otro ejemplo, actualmente en Argentina se debate acerca de cómo proceder ante discursos sobre la última dictadura (1976-1983) que se consideran negacionistas). Lo que sí se puede recomendar desde la teoría deliberativa es que el proceso para trazar ese límite sea objeto de debate público.

Dicho debate puede incluir a diferentes tipos de actores, desde la academia, pasando por la prensa, los partidos políticos y hasta la sociedad civil organizada y no organizada. Es decir que, en concreto, puede involucrar desde a una persona experta que participa mediante una columna de opinión en un periódico masivo, hasta una ciudadana de a pie que interviene en una conversación en la calle. Efectivamente, la concepción habermasiana de esfera pública abarca lo que se denomina la esfera pública informal e incluye, ya en sus primeras formulaciones, hasta conversaciones ocasionales (Habermas, 1974), y privilegia el lenguaje ordinario como puente hermenéutico (Habermas, 2000).

En cuanto a los efectos de este proceso en la esfera pública en relación a la tolerancia, serían técnicamente dos. Si una sociedad amplía el ámbito de las creencias tolerables, se asemejaría al tipo de sociedad representada en la figura 3, es decir, más tolerante. Si por el contrario, amplía el ámbito de las creencias y prácticas intolerables, se acercaría al tipo graficado en la figura 4, es decir, una sociedad polarizada.

En un sentido más general, si ésta ampliación del ámbito de creencias y prácticas tolerables se diera mediante la revitalización del debate público, en un círculo virtuoso de retroalimentación, podría hipotéticamente tener como efecto que ese carácter semipúblico que detecta Habermas evolucionara hacia un carácter más público; si esto se diera, representaría otra nueva transformación de la esfera pública, esta vez de signo positivo.

Lista de referencias

Carriquiry, A. (2019). De la «esfera pública plebeya» a las esferas públicas en plural. *Revista Encuentros Latinoamericanos*, III (2), 72-97.

Carriquiry, A. (2022a). Pensar con Habermas, después de Habermas: el rol de la prensa en la esfera pública (digital). *Revista Sistema*, 263, 49-64.

Carriquiry, A. (2022b). Jürgen Habermas y lo privado vuelto al público, en la esfera pública original y en la esfera pública digital. *Ideas y Valores*, 71 (180), 123-146.

Benhabib, S. (2018, 9 de octubre). Below the Asphalt Lies the Beach. *Boston Review.*

Block, H. y Riesewieck, M. (2018). *The cleaners*. [Documental]. Prod. Gebrüder Beetz Filmproduktion Köln GmbH & Co asociado a WDR, Grifa Filmes, NDR, RBB, BBC, VPRO and PlayTV.

Castells, M. (2008). The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. *The Annals of the American Academy of Political and Social Science*, 616 (1), 78-93.

Celeste, E. (2018). Terms of service and bills of rights: new mechanisms of constitutionalisation in the social media environment? *International Review of Law, Computers & Technology.*

Cohen, J. L. y Arato, A. (1992). *Civil Society and Political Theory*. Cambridge, MA: MIT Press.

Dahlberg, L. (1998). Cyberspace and the public sphere: Exploring the democratic potential of the net». *Convergence:The International Journal of Research into NewMedia Technologies*, 4 (1), 70-84.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, 22 (2), 147-162.

Dean, J. (2003). Why the net is not a public sphere. *Constellations,* 10(1).

Elster, J. (2000). «Arguing and Bargaining in Two Constituent Assemblies». University of Pennsylvania Journal of Constitutional Law, 2:2, 345-421.

Forst, R. (2013). *Toleration in Conflict. Past and Present,* Cambridge: Cambridge University Press.

Forst, R. (2014). Toleration and Democracy. *Journal of Social Philosophy, 45*(1), 65-75.

Forst, R. (2017). «Toleration» en Zalta, E. (Ed.), *The Stanford Encyclopedia of Philosophy*.

Feenberg, A. y Barney, D. (Eds.) (2004). *Community in the Digital Age*. Rowman & Littlefield.

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, 25/26, 56-80.

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

Gill, L., Redeker, D. y Gasser, U. (2015). Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights. *Berkman Klein Center for Internet & Society Research Publication*, 2015-15.

Habermas, J. (1953). Mit Heidegger gegen Heidegger denken: Zur Veröffentlichung von Vorlesungen aus dem Jahre 1935. *Frankfurt Allgemeine Zeitung*, 25/7.

Habermas, J. (1997). *Historia y crítica de la opinión pública: La transformación estructural de la vida pública* (5.ª ed.), Barcelona: Gustavo Gilli. (Obra original publicada en 1962).

Habermas, J. (1974). The Public Sphere: An Encyclopedia Article. *New German Critique*, 3, 49-55. (Obra original publicada en 1964).

Habermas, J. (1999). *Teoría de la acción comunicativa*, Madrid: Santillana. (Obra original publicada en 1981).

Habermas, J. (2000). *Facticidad y validez*, Madrid: Trotta. (Obra original publicada en 1992).

Habermas, J. (2006). «Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research». *Communication Theory*, 16(4), 411-426.

Habermas, J. (2007). How to save the quality press? *Signandsight.com*, 25/6/2007.

Habermas, J. (2014). «Internet and Public Sphere: What the Web Can't Do» / Entrevistado por Markus Schwering. *Reset: Dialogues on Civilizations*, 24/7/2014.

Habermas, J. (2016). «Für eine demokratische Polarisierung» / Entrevista. *Blätter für deutsche und internationale Politik,* noviembre, 35-42.

Habermas, J. (2016b). «The Cost and Challenge of the Eurozone Debt Crisis» / Entrevistado por Stuart Jeffries. *The Financial Times Weekend Magazine*, 1/5/2016.

Habermas, J. (2022). *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Berlín: Suhrkamp Verlag.

Habermas, J. (2018). «Jürgen Habermas: "Por Dios, nada de gobernantes filósofos!"» / Entrevistado por Borja Hermoso. *El País Semanal*. 25/4/2018.

Hawtin, D. (2011). «Cartas y principios de internet: tendencias y reflexiones». *Monitor Mundial sobre el Impacto de la Información. 2011*.

Iglesias Keller, C. (2018, 4 de octubre). Could fake news annul the Brazilian elections? *Digital Society Blog*. https://www.hiig.de/en/could-fake-news-annul-the-brazilian-elections/

Ilves, T. H. (2018, 25 de enero). Is Social Media Good or Bad For Democracy? *Facebook Newsroom*. https://newsroom.fb.com/news/2018/01/ilves-democracy/

Ipar, E. (2021). Habermas y el neoliberalismo. *Valenciana*, 13 (27), 223-249.

Kellner, D. (2000). «Habermas, the public sphere, and democracy: A critical intervention» en Hahn, L. E. (Ed.), *Perspectives on Habermas* (pp. 259-288). Chicago, IL: Open Court Press.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Nueva York: Penguin.

Schäfer, M. (2015). «Digital Public Sphere» en Mazzoleni, G. (Ed.), *The International Encyclopedia of Political Communication* (pp.322-328), Nueva York: Wiley and Sons.

Sunstein, C. (2001). *Republic.com*, Princeton: Princeton University Press.

Sunstein, C. (2007). *Republic.com 2.0. The revenge of the blogs,* Princeton: Princeton University Press.

Sunstein, C. (2017). *#Republic. Divided democracy in the age of social media,* Princeton: Princeton University Press.

Sunstein, C. (2018, 22 de enero). Is Social Media Good or Bad for Democracy? *Facebook Newsroom*. https://newsroom.fb.com/news/2018/01/sunstein-democracy/.

The Conversation (2024). https://theconversation.com/global

Velasco, J. C. (2019). «Migrants and Refugees» en Allen, A. y Mendieta, E. (Eds.), *The Cambridge Habermas Lexicon* (pp. 30-35). Nueva York: Cambridge University Press.

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 55-70
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.610891

Absolute Freedom of Speech and Social Media: Deconstructing the Argument of Individual Self-Realization

La libertad de expresión absoluta y las redes sociales: Deconstruyendo el argumento de la autorrealización individual

KEBERSON BRESOLIN[*]

Abstract: The paper challenges the absolute conception of freedom of speech as an unconditional means for individual self-realization. Firstly, it discusses the positions of Scanlon and Redish, revealing the inherent vulnerabilities in their arguments. Subsequently, it argues against the view of unlimited freedom of speech as fundamental to self-realization. Finally, even if one were to accept the premise of self-realization as an axiom, social media would not qualify as suitable arenas for its actualization, given their inability to replicate the fundamental characteristics of a public sphere that favors open, plural, and rational debate.
Keywords: Self-Fulfillment, Autonomy, Scanlon, Social Media, Public Sphere, Habermas.

Resumen: El artículo cuestiona la concepción absoluta de la libertad de expresión como medio incondicional para la autorrealización individual. Inicialmente, se discute la posición de Scanlon y Redish, revelando las vulnerabilidades inherentes a sus argumentos. A continuación, se argumenta en contra de la visión de una libertad de expresión ilimitada como esencial para la autorrealización. Finalmente, aun aceptando la premisa de la autorrealización como axioma, las redes sociales no se calificarían como arenas adecuadas para su efectuación, dada su incapacidad para replicar las características fundamentales de una esfera pública que favorezca el debate abierto, plural y racional.
Palabras clave: Autorrealización, Autonomía, Scanlon, Redes Sociales, Esfera Pública, Habermas.

Recibido: 04/04/2024. Aceptado: 24/06/2024.

* Dr. Keberson Bresolin holds a Ph.D. in Philosophy from the *Pontifical Catholic University* of Rio Grande do Sul, with a research period at *Eberhard Karls Universität Tübingen* under the supervision of the esteemed Prof. Dr. Dr. h.c. Otfried Höffe. He completed his postdoctoral studies at the University of Tübingen with a scholarship from the *Alexander von Humboldt Foundation*. Dr. Bresolin is currently a professor at the Graduate Program in Philosophy at the *Federal University of Pelotas* (UFPel) and a collaborator in the Graduate Program in Philosophy at the *Federal University of Rondônia*. His research focuses on Kantian philosophy, social medias, political philosophy, and contemporary intersections between philosophy of law and politics. Notable publications include 'Federalism' in *Revista Opinião Filosófica* (2024), and 'Freedom of Expression, Public Sphere, Digital Platforms, and Social Media' in *ETHIC@* (2023). E-mail: keberson.bresolin@ufpel.edu.com

1. Preliminary Considerations

Freedom of speech is commonly defined as the inalienable right of every individual to express their opinions, ideas, and thoughts, free from fear of retaliation or censorship by government entities, society, or other individuals. This concept has its roots deeply intertwined with the evolution of individual rights and the strengthening of democracy. In contemporary times, it is observed that some groups advocate for freedom of speech in an absolute manner, without, however, basing their claims on robust arguments that give it a convincing legal or philosophical foundation (Bresolin, 2023a: 764).

In this paper, I will explore an argument often mobilized in defense of the concept of absolute and unrestricted freedom of speech, specifically, the principle of individual self-realization. This argument is fundamentally distinct from the Millian conception of the «Marketplace of Ideas» (Bresolin, 2023b: 469), positing that any restrictions on freedom of speech significantly compromise the development of individuals' capabilities and autonomy. I will argue that this perspective faces various significant problems that challenge its consistency and theoretical sustainability.

After deconstructing the argument in favor of absolute freedom of speech, I intend to demonstrate that the environment of social media, often considered a space for the manifestation of free expression, does not constitute a legitimate extension of the public sphere. Due to the control exercised by large technology corporations, social media do not promote an open, free, and plural debate. On the contrary, through algorithms that create «filter bubbles», these platforms facilitate the self-confirmation of pre-existing ideas and exacerbate polarization. This mechanism limits exposure to divergent perspectives and, consequently, restricts the possibility of a genuinely democratic and constructive dialogue in the digital sphere.

In this context, even if we were to adopt the argument of individual self-realization as a justification for the defense of absolute freedom of speech, such freedom would not find practical applicability on social media platforms. These platforms do not favor the autonomy and self-development of the individual; on the contrary, the operational dynamics of social media, driven by algorithms that shape the user experience, tend to restrict the spectrum of ideas and information accessible. This limitation directly interferes with the process of critical formation and the capacity for self-development of the individual, compromising the fundamental basis of the self-realization argument.

2. Individual Self-Fulfillment

While the argument of freedom of speech as a discovery of truth presents itself as consequentialist, as a means to an end, freedom of speech as an aspect of Self-Fulfillment carries an intrinsic value, particularly according to Scanlon's conception (1972). Thus, this view understands freedom of speech as a fundamental aspect of the individual's right to self-development and realization, that is, a right as an «intrinsic and independent good» (Barendt, 2007: 13). Therefore, any form of censorship of freedom of speech becomes an obstacle to the growth and personality of the individual. Redish advocates that any external judgment claiming that a certain expression promotes Self-Fulfillment more than another is

a violation of the individual's free will, as recognizing this is fundamental to the principle of Self-Fulfillment (Redish, 1982: 592). This author champions the value of Self-Fulfillment, emphasizing freedom of speech as fundamental for individual development and self-governance. He identifies two essential components in Self-Fulfillment: the development of individual skills and capabilities (self-development) and control over one's own destiny through impactful life decisions (self-governance).

In this manner, Redish argues that freedom of speech directly promotes self-development, as free expression is an essential tool for personal growth, enabling people to explore and express their ideas, which in turn contributes to the development of their cognitive and emotional skills. The ability to express oneself freely is seen as fundamental to personal evolution, a central aspect of Self-Fulfillment. Freedom of speech promotes self-governance only indirectly, to the extent that it provides a free flow of information and opinions that guide people in their decisions (Baker, 1982: 668).

Redish considers Self-Fulfillment central to democracy, arguing that it entails the protection of freedom of speech, given that democracy promotes values of self-governance and self-development. Democracy is nothing more than a means for the development and Self-Fulfillment of individuals. According to Redish, proponents of the freedom of speech argument for the sake of democracy, like Meiklejohn (1948), have confused a means of obtaining the final value with the value itself. This implies that, if an individual has the opportunity to control their destiny, it is essential that they have access to all pertinent information in order to collaborate in efficient decision-making, thus directly impacting their life. The principle of self-rule is termed an «intrinsic» value, as it is achieved through the very existence of a democratic system (Redish, 1982: 601-621).

In this sense, given that people make choices daily that reverberate in their lives, from those that seem relevant to those considered trivial, it is possible to infer that any opinion or information, no matter how insignificant it may seem, may affect such decisions at some point.

The secondary value of a democratic system is designated as instrumental, since it consists of a purpose for which the democratic system is designed to lead, in contrast to a goal that is achieved by definition — intrinsic — through the adoption of the democratic system itself, that is, the promotion of the development of human faculties. In this way, Redish concludes:

> My thesis is that: (1) although the democratic process is a means of achieving both the intrinsic and instrumental values, it is only one means of doing so; (2) both values (which, as noted previously, may be grouped under the broader heading of self-realization) may be achieved by and for individuals in countless nonpolitical, and often wholly private, activities; and (3) the concept of free speech facilitates the development of these values by directly fostering the instrumental value and indirectly fostering the intrinsic value. Free speech fosters the former goal *directly* in that the very exercise of one's freedom to speak, write, create, appreciate, or learn represents a use, and therefore a development, of an individual's uniquely human faculties. It fosters the latter value *indirectly* because the very exercise of one's right of free speech does not in itself constitute an exercise of one's ability to make life-affecting decisions as much as it facilitates the making of such decisions (Redish, 1982: 603-604).

Democracy, as a form of government, affords individuals the opportunity to achieve Self-Fulfillment, through the refinement of skills and abilities, as well as the autonomy inherent in the right to govern one's own existence. From this perspective, freedom of speech is a fundamental pillar for the valorization of the process of human Self-Fulfillment. Thus, the protection of freedom of speech concerns not just political judgments, but rather to promote the broader values that the democratic system was designed to foster.

However, unless grounded in arguments that demonstrate the particular relevance of expression, the argumentation in favor of the principle of freedom of speech, as a means to Self-Fulfillment, becomes difficult to distinguish from the more comprehensive claims of libertarianism, which defend the right to do anything considered an integral part of the individual's personality.

From the same perspective, it is plausible to question why freedom of speech holds a prominent position in the pursuit of individual Self-Fulfillment. It cannot be unequivocally stated that unlimited freedom of speech inevitably triggers personal challenges or that it is rooted in more fundamental human needs and desires than other necessities, such as education and adequate housing.

To deepen the analysis of the previous statement, one could use Brazil as an example, where 31.6% of the population finds themselves in a condition of poverty (Gomes, 2024). In this context, the concept of a minimum existential, which ensures basic social rights — such as health, food, and education — is undoubtedly crucial. These rights are fundamental to guarantee the minimum vital conditions necessary for human subsistence and freedom of action (Espinoza, 2017: 110).

Given the complexity of establishing criteria that prioritize freedom of speech, there is a shift towards a libertarian orientation, in which freedom is considered the most precious and fundamental good. This approach postulates that, beyond any other valuation, individual freedom should be placed on a pedestal, suggesting that restrictions on freedom of speech should be exceptional and justifiable only in cases where there is a direct and concrete harm to others. Embedded in this perspective is a naive premise about the capacity and intrinsic commitment of individuals to exercise their freedom in a conscious and respectful manner, especially on social media, in a way to harmonize their own expressions with the freedoms of choices of others.

In the same vein, freedom of speech is intertwined with other fundamental freedoms that denote an aspect inherent to the human condition, such as religious freedom, freedom of thought, and freedom of conscience. However, unlike the latter, freedom of speech can harm others when it is exercised, for example, by damaging a person's reputation or infringing upon privacy and intellectual property rights (Barendt, 2007: 13-14).

Similarly, Baker, in analyzing Redish's claims, brings to light profound questions about the essence of Self-Fulfillment and the role of freedom of speech in democratic society. Redish argues that freedom of speech directly promotes self-development, but only indirectly self-rule, suggesting that «speech directly fosters self-development but only indirectly fosters self-rule» (Baker, 1982: 658). This distinction is crucial for understanding Redish's approach, in which he views freedom of speech as a tool for the dissemination of information that, in turn, would enable self-governance indirectly. Baker, however, questions this separa-

tion, arguing that such a view underestimates the direct importance of freedom of speech as a means of exercising autonomy and actively participating in democratic governance.

A significant point of controversy between the two theorists emerges in the interpretation of democracy's role in Self-Fulfillment. Baker criticizes Redish's assumption that democratic acceptance necessarily implies valuing self-development, stating that «Redish fails to show that our acceptance of democracy logically implies acceptance of the self-development value or that this value underlies the First Amendment» (Baker, 1982: 660). This critique points to a broader and less restrictive conception of freedom of speech, which is not limited to promoting self-development, but encompasses a wider range of expressive activities essential to democracy.

Baker also challenges the notion that democracy is a requirement for self-development, arguing that the relationship proposed by Redish between these two concepts is not as direct as suggested. According to Baker, «although democracy may further the ‹development of the individual's human faculties›, a concern with self-development does not, in any obvious way, require a democratic political order» (Baker, 1982: 660). Baker suggests that other political systems could equally promote self-development, questioning the exclusive link made by Redish between democracy and Self-Fulfillment.

Finally, Baker argues that any justification for the constitutional protection of freedom of speech based on the instrumental contribution of speech to self-rule is insufficient. He raises concerns about the possibility that additional speech may, in fact, harm self-rule, contributing to information overload, presenting a distorted or ideologically unbalanced perspective, or promoting simplistic thinking. Baker suggests that the protection of freedom of speech should be grounded in considerations beyond its indirect or instrumental contribution to self-rule, focusing on freedom of speech as a constitutive aspect of self-rule itself (Baker, 1982: 663-664).

3. Scanlon's Argument

In turn, Scanlon defends freedom of speech on the premise that «the powers of a state are limited to those that citizens could recognize while still regarding themselves as equal, autonomous, rational agents» (Scanlon, 1972: 215). Although Scanlon referred to his position as a «natural extension» (Scanlon, 1972: 213) of Chapter II of Mill's *On Liberty* (2015), his argument significantly diverges from Mill's consequence-based argument and, instead, finds its argumentative roots in Kant and Rawls. Scanlon's argument is not grounded in claims about the consequences of different policies, but rather aims to offer an alternative to the conventional view based on the premise of rational and autonomous agents.

Although Scanlon makes it clear that his proposal aligns with the Millian principle, his theory does not present itself as a consequentialist theory, but one founded on rights. As such, it does not argue that truth will necessarily be attained. Through the discussion of examples that could restrict freedom of speech, Scanlon presents his Millian principle:

> There are certain harms which, al-though they would not occur but for certain acts of expression, nonetheless cannot be taken as part of a justification for legal restrictions on these acts. These harms are:

> (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of ex-pression;
>
> (b) harmful consequences of acts performed as a result of those acts of expression, where the connection between the acts of expression and the subsequent harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing (Scanlon, 1972: 213).

Acording to Scanlon, Mill's principle is an absolute criterion within its sphere, aimed at entirely excluding «certain justifications for legal restrictions on acts of expression», and thus, should be «the basic principle of freedom of speech» (Scanlon, 1972: 214). In both cases, the harmful outcome was not the deliberate intention of the author of the act of expression. In one of his examples of limiting freedom of speech, Scanlon posits that a person, through an expressive act, may contribute to the generation of a harmful act committed by another. In certain situations, the negative effects resulting from the second act may justify classifying the first as a crime (an order, for example) (Scanlon, 1972: 2011).

In this aspect, Brison (1998), on the other hand, critiques this view for not fully recognizing the social and psychological impact of hate speech. She argues that hate speech not only propagates false and harmful beliefs about individuals or groups but also generates real and tangible harms, such as diminished self-esteem and the perpetuation of systems of discrimination. Brison emphasizes that these harms, both in the formation of false beliefs (category a) and in the harmful actions resulting from these beliefs (category b), are sufficiently serious to justify restrictions on hate speech. She contends that protection against these harms is necessary to preserve human dignity and social equality, values that also underpin freedom of speech (Brison, 1998: 323).

The concept of autonomy is the foundation of Scanlon's theory of freedom of speech. He regards individual autonomy as the locus of human realization, with the defense of freedom of speech being indispensable in this process. He states that his concept of autonomy does not require the prerequisites of the Kantian concept of autonomy, in such a way that he advocates that «to be autonomous in my sense is quite consistent with being subject to coercion in relation to one's own actions» (Scanlon, 1972: 216). A very weak concept of autonomy, in Scanlon's view, is sufficient to establish the framework from which governmental authority is prevented from performing any kind of intrusion. In this sense, he argues that,

> To regard himself as autonomous in the sense I have in mind a person must see himself as sovereign in deciding what to believe and in weighing competing reasons for action. He must apply to these tasks his own canons of rationality, and must recognize the need to defend his beliefs and decisions in accordance with these canons. This does not mean, of course, that he must be perfectly rational, even by his own standard of rationality, or that his standard of rationality must be exactly ours. Obviously, the content of this notion of autonomy will vary according to the range of variation we are willing to allow in canons of rational decision. If just anything counts as such a canon then the requirements, I have mentioned will become mere tautologies: an autonomous man believes what he believes and decides to do what he decides to do (Scanlon, 1972: 215).

Scanlon explicitly states that he will not describe a set of limits on what he considers to be canons of rationality. According to him, the most important consideration is that an autonomous individual cannot simply accept, uncritically, the judgments of others regarding their conduct and beliefs. It is possible for them to accept external evaluation; however, it is necessary that they have the autonomy to analyze the probative value of the judgments presented, as well as to justify autonomous and independent reasons that demonstrate the veracity of these judgments, so that they can weigh them against contrary evidence and establish their own autonomous judgment (Scanlon, 1972: 216).

Arguing in favor of hate speech freedom and contesting restrictions on this practice, Nagel (1995) has referenced a similar conception of autonomy. For him, the condition of being an independent thinking being demands that the expression of thought and feelings should be, primarily, an individual responsibility, limited only by clearly necessary restrictions to prevent serious harm distinct from the expression itself. This quality establishes a moral recognition that each individual holds their own opinions, and that the possibility of impeding their right to expression is a violation of their integrity. As an aspect of status, freedom of speech intertwines with freedom of thought, since to suppress it also means to repress a fundamental aspect of the shared cognitive process through which the mind can develop freely, as we work, while we think, as participants in a collective endeavor (Nagel, 1995: 96).

In this conception, the worst consequence lies in the censorship of dissenting opinions due to the risk of persuasion they may exert on people, thus failing to support established orthodoxy. Such an attitude is described by Nagel as epistemological stupidity since it constitutes the ultimate insult not only to the dissenters but also to us, as the potential public, insolently suggesting our incapacity to make independent decisions. One could not be jailed or fined for denying, for example, that the Holocaust occurred, or for selling books that deny it, or for running a mail-order business selling Nazi medals.

4. Critical Considerations on the Self-Fulfillment Thesis

That said, Barendt notes that the subject has the right to listen to different viewpoints and consider acting upon them, even if such a procedure might be detrimental to society, acknowledging, however, that certain restrictions may be applied. In contrast, unlike other approaches to the Self-Fulfillment argument, Scanlon focuses on the rights and interests of those who are recipients of the communication (Barendt, 2007: 16). However, Scanlon's Millian principle rests on the limitation of governmental authority and not on a right of individuals or, according to Nagel, as a general moral right—a universal human right,—which, as Brison rightly observed, Nagel does not provide an explanation for why the right to freedom of speech should be considered as such (Brison, 1998: 327).

It's important to highlight how Scanlon transitions from the idea of rational agents to his Millian principle. Two arguments are offered. The first is based on an appeal to our intuitions about agency and responsibility, namely, Scanlon believes that a rational adult assumes full responsibility for their actions and decisions to act. By following their beliefs and judgments, deemed sufficient to justify their action, they cannot blame the people who

provided the reasons for acting for the harm caused. Transferring this responsibility would deny the agent's own autonomy and rationality.

The second argument advocates that, from a perspective where a group of rational and autonomous citizens finds themselves in an original position similar to Rawls's, it is feasible to assert that such individuals would not grant the State the authority to determine the type of arguments that could be heard once the veil of ignorance is lifted. Granting such authority to the government (or any other entity) would be an affront to the autonomy and rationality of these agents, undermining their position as free and conscious decision-makers (Amdur, 1980: 290-293).

The conception of freedom of speech as Self-Fulfillment in the variant presented by Scanlon has faced numerous critiques. Amdur concludes, upon examining various real and hypothetical cases, that individuals who provide persuasive reasons for harmful actions are also morally responsible for the resulting damages. Intuitions about moral responsibility do not support the Millian principle; on the contrary, they raise serious doubts about whether the Millian principle could be correct. Amdur discusses whether our intuitions about legal responsibility support the Millian principle, even if our intuitions about moral responsibility do not. The author suggests that we do not have clear intuitions about legal responsibility, but if we do, they likely reflect moral responsibility. However, there may be reasons to deviate from these intuitions, such as the difficulty in identifying morally responsible individuals. Ultimately, the author concludes that our intuitions about legal responsibility do not support Mill's principle (Amdur, 1980: 297).

Even stating that the Millian principle does not work, Amdur asserts that, if it did work, there would be the following problem: the Millian principle proves too much because, if we accept the claims about responsibility, it is not clear why the State can legitimately restrict the expression of acts that Scanlon is willing to restrict. The formulation invites the question of why the contribution to the genesis of Jones's action to rob a bank, for example, made by the act of expression is also not «superseded by the agent's own judgment» (Scanlon, 1972: 212) when he chooses to manufacture and use nerve gas after reading Smith's formula. Jones must decide to manufacture and use the nerve gas with the same certainty that he must decide to rob the bank (Amdur, 1980: 297).

Following the influence of Rawls's original position, Amdur points out that Scanlon did not consider all the configurations of the parties' decisions. In this sense, autonomous citizens would consider both their own rights to speak and hear different viewpoints, as well as the potential harms caused by acts of expression. They might reach an agreement prohibiting the State from interfering in expression based on content, but allowing an exception for acts that cause serious harm. The author suggests that citizens would not demand a principle as strict as the Millian principle (Amdur, 1980: 299).

Amdur's last critique is directed at a certain arbitrariness in what the Millian principle might cover. Scanlon encounters difficulties in applying the Millian principle to cases that seem relevant. During the discussion on the various ways in which acts of expression can cause harm, he drafts the following passage: «Another way in which an act of expression can harm a person is by causing others to form an adverse opinion about her or making her an object of public ridicule. Obvious examples of this are defamation and interference with the right to a fair trial» (Scanlon, 1972: 211).

However, Scanlon's belief that the State can restrict certain types of expression and not others is at least questionable. Scanlon argues that, for example, if A's statements lead B to form an adverse opinion about C, the State can intervene, but if A's statements lead to B murdering C, the State cannot intervene. This view is curious, as it is not clear why the State should be allowed to restrict defamation but not incitement (Amdur, 1980: 300). There is no plausible explanation for the differentiation, so that if the Millian principle prohibits restrictions on incitement, it should also prohibit restrictions on defamation.

It is still possible to question again the presupposition of Scanlonian autonomy, as the theory struggles to respond to the accusation that numerous people are often factually incapable of exercising their autonomy. It is assumed, therefore, that many of them are unable to consider all the viewpoints and arguments presented to them. Consequently, absolute freedom of speech may trigger unwise and highly dangerous choices, such as voting for candidates who self-proclaim as heralds of freedom but violate the principles of pluralism and tolerance.

Furthermore, Scanlon's theory of freedom of speech and autonomy encounters particularly challenging terrain in the context of social media. The Scanlonian premise of autonomy presupposes that individuals are capable of exercising their freedom in a rational and informed manner, considering a plurality of viewpoints and arguments. However, the characteristics of social media, such as the echo chamber phenomenon and its distinct nature from a traditional public sphere, exacerbate the inability of many users to actually exercise this proposed autonomy.

As I will argue further, social media do not function as a public sphere in the Habermasian sense, where rational critical discourse among informed citizens could prevail. Instead, these platforms tend to segment users into niches or echo chambers, within which they are predominantly exposed to opinions and information that reinforce their pre-existing beliefs. Such a structure not only makes it difficult to be exposed to a diversity of arguments, as suggested by Scanlon as essential for autonomy, but also amplifies the spread of misinformation and polarizing discourse.

Moreover, the issue of being unfit to exercise autonomy becomes more pressing in the social media environment. Information overload, the speed at which news spreads, and the partial anonymity offered by these platforms can encourage unreflective and impulsive decisions. This directly contradicts the idea of a deliberative and well-informed autonomy, crucial to the Scanlonian theory.

Therefore, when considering the impact of social media on individuals' ability to effectively exercise autonomy, it becomes evident that Scanlon's theory of freedom of speech seems insufficient to deal with the contemporary challenges posed by these platforms.

5. Social Media Do Not Promote Individual Self-Fulfillment

Even if the premise of Self-Fulfillment were accepted as an axiom, social media would not be configured as appropriate arenas for its realization, due to their inability to replicate the essential characteristics of a public sphere conducive to fostering an open, diverse, and rational debate.

That said, according to *Demandsage* (Shewale, 2024), the latest data shows that 5.17 billion people use social media in 2024, which equates to 63.82% of the world's population. The research further projects that the number of users is expected to reach 5.85 billion (Shewale, 2024). From the context presented and the data provided, it is evident that social media have established themselves as an inescapable reality in contemporary society. Therefore, these platforms have transcended their initial role as spaces for social interaction to become primary sites of freedom of expression manifestation.

On one hand, this scenario reflects a profound transformation in the way freedom of speech is exercised globally. By providing an open and accessible platform for the dissemination of ideas, opinions, and information, social media democratize expression in an unprecedented manner. Individuals from different parts of the world, with varying levels of access to resources and traditional media platforms, find in social media a means to express their views, participate in public debates, and influence discourses on a global scale.

From this perspective, Shirky (2011), in addressing the dangers of freedom on the internet, argues that two perspectives can be developed, namely, instrumental and environmental. The instrumental perspective emphasizes the promotion of freedom of access as an essential pillar of this approach. This focus highlights the importance of unrestricted access to global information and the ability of citizens to generate public media in countries under authoritarian regimes, in addition to freedom of speech for activists and the use of instant messaging without interference.

In contrast, the environmental perspective offers an alternative view, conceiving new media as facilitators of citizen participation and strengtheners of individual and collective freedoms. This view maintains that, akin to previous innovations such as the printing press and postal service, modern digital technologies have the potential to foster a robust public sphere and a vibrant civil society. It emphasizes the ability of dissident movements to use any available means to articulate their views and coordinate their actions, thus challenging authoritarian governments that fear unrestricted communication among their citizens. The environmental perspective criticizes the instrumental approach for its difficulty in grasping local conditions of dissent and the risk of compromising the integrity of peaceful opposition through external support. Instead, it proposes a long-term view of social media as tools that lay the groundwork for sustainable democratic transformations, arguing that positive changes follow, rather than precede, the development of an engaged and informed public sphere (Shirky, 2011: 2-4).

With a more enthusiastic and not entirely incorrect view, Loader and Mercea argue that social media have a disruptive effect on dominant discourse: «Equipped with social media, citizens no longer have to be passive consumers of political party propaganda, government spin, or mass media news but are instead actually enabled to challenge discourses, share alternative perspectives, and publish their own opinions» (Loader; Mercea, 2011: 759). Indeed, social media equip citizens with tools to question narratives controlled by powerful entities, promoting an environment where multiple voices can be heard and considered. Such a phenomenon is evidence of the democratizing potential of social media, which challenges the monopoly of media production and its dissemination by state and commercial institutions.

Firstly, when considering the democratizing potential of social media, one cannot neglect the propensity of these same platforms for the creation and strengthening of echo chambers (Samaržija, 2023). These echo chambers are virtual spaces where ideas and beliefs are amplified through repetition within a closed community, often isolating its members from divergent or contradictory opinions. This phenomenon results in heightened polarization, where dissent and critical debate are replaced by an illusory consensus, often built upon unexamined or even false premises.

Moreover, social media platforms are designed to maximize user retention and engagement, which often translates into the promotion of content that provokes strong emotional reactions, rather than balanced information or diverse perspectives. Such dynamics can inadvertently favor the proliferation of dominant discourses, instead of challenging them, as polarizing or sensationalist narratives tend to receive greater visibility and dissemination.

Another crucial aspect to consider is the role of algorithms that govern what is seen or not by users on social media. These algorithms, often opaque and devoid of accountability, can intensify exposure to homogeneous viewpoints and filter out information that contradicts the user's pre-existing beliefs, thus reinforcing echo chambers and limiting the potential for genuinely diverse and constructive dialogue (Samaržija, 2023: 72-74).

In this regard, while social media undoubtedly possess the potential to challenge the monopoly of media production and promote a broader spectrum of voices, the practical reality of their operation reveals a complexity that can, paradoxically, reinforce dominant discourses and restrict the diversity of perspectives.

Loader and Mercea (2011) highlight the transformative potential of social media in reconfiguring power relations in the sphere of communication. By arguing that «by facilitating social networking and "user-centred innovation", citizens are said to be able to challenge the monopoly control of media production and dissemination by state and commercial institutions» (2011: 759), they point to a paradigm shift in which user-centered innovation and social networks enable individuals to question and challenge the traditionally monopolistic dominance of state and commercial institutions over media production and dissemination. This process not only evidences the decentralization of media power but also promotes broader and more diverse participation in the construction of public discourse, marking a significant step towards a more effective democratization of communication in contemporary society.

On the other hand, a less optimistic view of the role of social media as a public space suggests that, from the perspective of this new arena of expression, a distinctive element emerges, namely: digital communication platforms represent entities not traditionally framed as media. These platforms forgo the productive function of journalistic mediation and programming, attributes inherent to classic media, thus reshaping the communicational paradigm previously prevalent in the public sphere. Such platforms grant all potential users the ability to emerge as autonomous authors, endowed with comparable rights. The innovative nature of these technological infrastructures lies in providing their users unlimited opportunities for digital interconnection, functioning as blank slates (*leere Schrifttafeln*) for the inscription of their own communicative contents. However, they choose not to assume editorial responsibility for the content they distribute, in contrast to what is observed with news services or classic editors, such as the written press, radio,

or television, where the communicative content is produced professionally and subject to editorial filtering (Habermas, 2022: 43).

According to Habermas, it is possible to identify two significant impacts resulting from the structural transformation in the public sphere, spurred by the advent of a new pattern of communication. Initially, the universalist aspiration of the bourgeois public for an egalitarian inclusion of all citizens seemed finally achievable through the advent of new media. These promised to emancipate users from the traditional passive role of mere recipients, limited to selecting from a restricted range of programs, granting each individual the opportunity to express themselves within an anarchic exchange of spontaneous opinions. However, this potential, simultaneously anti-authoritarian and egalitarian, ends up morphing into a libertarian character, characteristic of digital corporations that dominate the global scene.

Within this context, the new media provides a stage both for extremist right-wing networks and for the intrepid Belarusian women who stand firm in their protests against Lukashenko. The self-empowerment (*Selbstermächtigung*) afforded to users of digital media constitutes one side of the coin; the reverse is the burden represented by the liberation from the editorial curation characteristic of traditional media, before users acquired the necessary competence to adequately handle the resources of the new media (Habermas, 2022: 45-46). From this perspective, the natural and spontaneous generation of vast communication networks around certain themes or personalities can lead to fragmentation, as such connections tend to group themselves into communicational circuits that isolate each other dogmatically.

The research conducted by Krause, Norris, and Flinchum (2017) offers a penetrating look at the dynamics between social media and the Habermasian concept of the public sphere. Through a detailed study, they reveal a reality far from the democratic idealization of social media as a revitalized space for rational public debate.. Three crucial arguments stand out that, together, underpin the perspective that social media do not meet the necessary criteria to be considered part of the public sphere.

Firstly, the prevalence of a *lack of civil discourse* on social media is alarming. The public sphere, as idealized by Habermas, is a domain of social interaction where individuals can discuss and deliberate on matters of common interest in a rational and respectful manner. Habermas emphasizes that what is considered public (*Öffentlich*) is characterized by being «accessible to all» (*allen zugänglich sind*) (Habermas, 2001: 54). According to Habermas, the public sphere constitutes a domain in which subjects have the possibility to group together and participate in a civic debate grounded in reason, aimed at the collective interests of society (*zivilen Aufgaben einer öffentlich räsonierenden Gesellschaft*) (Habermas, 2001, 116).

However, the data collected by Krause, Norris, and Flinchum (2017) indicate that social media are marked by a toxic communication environment, where personal attacks, disrespect, and polarization replace constructive debate. This scenario of virtual hostility inhibits not only meaningful participation but also the possibility of reaching a consensus or mutual understanding on political and social issues. The absence of civil dialogue prevents social media from functioning as an authentic public sphere, where discourse can flow freely and constructively.

The second argument focuses on the *limitations to information access* and *participation imposed by surveillance* and the fear of *online harassment*. Surveillance, whether state, institutional or interpersonal, and the consequent self-censorship, act as restrictive forces that

shape online behavior. The fear of reprisals—professional, personal or social—leads to a retreat in the expression of opinions and participation in debates. These dynamics create barriers to access and the free exchange of information, vital components of the public sphere. Without the possibility of open and unrestricted discourse, social media fail to promote an environment conducive to the formation of an informed and active public opinion.

Lastly, the tendency toward the formation of «echo chambers» and self-censorship reinforces polarization instead of fostering inclusive dialogue. On social media, the selection of contacts and the personalization of content lead to limited exposure to divergent perspectives. This homogenization of discourse encourages the formation of isolated groups, within which opinions are reaffirmed without being challenged. This isolation directly contradicts the Habermasian principle of the public sphere, which presupposes interaction between different worldviews as a means to enrich democratic debate and strengthen the social fabric (Krause, Norris, and Flinchum, 2017: 8-14).

Habermas argues against the simplistic description of digital platforms as «vehicles of interconnected communicative content on any scale» because he considers them misleadingly neutral and impartial. The presumed neutrality is refuted by the operation of these platforms under the control of algorithms, exemplified by giants like Facebook, YouTube, Instagram, Twitter, and TikTok. These social networks, operated by corporations among the most globally valued due to their significant market value, follow capitalist logic. The profit of Big Techs primarily comes from the collection and sale of data for advertising or other commercial purposes. These data, generated as by-products of user interactions with the platforms, comprise personal information accumulated on the internet by users (Habermas, 2022: 53-54). Therefore, the notion of neutrality of these platforms is belied by the reality of their commercial practices focused on data exploitation.

Habermas also presents the thesis that digital platforms induce the formation of semi-public spheres (*Halböffentlichkeit*) and self-directed, which emerge spontaneously. These spheres distance themselves not only from the traditionally editorial or official public sphere but also from each other, promoting a dynamic of mutual and reflective confirmation of perceptions and pronouncements. This phenomenon favors the creation of fertile ground for the multiplication of narratives and viewpoints limited in their reach and diversity (Habermas, 2022: 58). Anticipating this observation, Sunstein, in his work *Infotopia* (2006), already expressed concern about so-called information cocoons – communicational spaces where echoes of individuals' own choices and preferences predominate, «communication universes in which we hear only what we choose and only what comforts and pleases us» (Sunstein, 2006: 9). This mechanism contributes to the deterioration of deliberative debate and reinforces existing prejudices. Similarly, Eli Pariser, in 2011, brought to light the notion of «filter bubbles», arguing that the effects of algorithmic filtering lead Internet users to receive information that resonates exclusively with their pre-existing interests. This process results in isolation about divergent views and can significantly limit individuals' freedom of choice on how to live. Pariser also warns about «informational determinism», where previous web interactions shape future content exposures, trapping users in a repetitive cycle of information they already know — «a web history you're doomed to repeat» (Pariser, 2011: 13-14).

Furthermore, in this context, Big Techs position themselves as the heralds of impartiality and freedom of expression by not producing, editing, or selecting content.[1] On the other hand, by creating new connections as «irresponsible» (*unverantwortliche*) mediators in the global network and initiating and intensifying discourses of unpredictable content – such as Fake News and Hate Speech — through the surprising acceleration of contacts, they profoundly alter the character of public communication (Habermas, 2022: 44).

The reflections of Habermas, Sunstein, and Pariser converge on a critical point that challenges the conception of social networks as effective spaces of the public sphere. Although social networks boast a remarkable potential to facilitate open dialogue and the democratization of information, the intrinsic dynamics of these digital platforms contradict the fundamental requirements for the constitution of a truly democratic and inclusive public sphere.

Concluding remarks

This discussion aimed to demonstrate the relationship between the argument for absolute freedom of speech through the defense of individual Self-Fulfillment and its connection with social media. The argument for absolute, unrestricted freedom of speech does not hold through the argument of individual Self-Fulfillment. Numerous criticisms of this thesis were listed, highlighting the argument of *libertarianism versus social responsibility* and *the universality of freedom of speech*, which question the premise that freedom of speech, seen as a vehicle for individual Self-Fulfillment, can be considered an absolute priority over other fundamental needs and rights. By placing freedom of speech on a pedestal, disregarding potential conflicts with equally important rights, such as human dignity, social equity, and protection against harmful forms of expression (e.g., hate speech and misinformation), this perspective ignores the social and psychological impact of certain expressions. Additionally, the unilateral emphasis on Self-Fulfillment through expression ignores complex socio-economic realities, where basic rights like health, education, and housing may take precedence in the hierarchy of individual and collective needs, especially in contexts of poverty and inequality.

Regarding the *relationship with democracy* and *Scanlon's principle of autonomy*, there is a fundamental critique of the assumption that freedom of speech, as a pillar of Self-Fulfillment, is essential to democracy and self-rule. This critique challenges the idea that democracy primarily serves as a means for individual Self-Fulfillment, suggesting that this reductionist view may neglect essential aspects of democratic participation and the balance between individual freedom and collective well-being. Moreover, while Scanlon's principle of autonomy seeks to establish freedom of speech on grounds of equality and autonomous rationality, critics argue that his approach does not adequately address the practical challenges posed by harmful discourse, underestimating the impact of expressions that can compromise human dignity, social cohesion, and the right not to be psychologically affected.

1 To corroborate this, it's enough to mention Google's campaign in Brazil against the "Fake News Bill." The platform displayed a message on the search engine's homepage, stating that the bill could "increase confusion about what is true or false" (Pinotti, 2023). It was clearly offering a biased view of the "Fake News Bill" to favor its corporate and economic interests.

Considering this, the defense of the thesis that social media do not constitute a true public sphere due to limitations imposed by echo chambers and algorithms promoting polarization and homogenization of the debate, contrary to the principles of diversity of opinions and rational deliberation characteristic of a public sphere, was advanced. Even if all the severe objections directed at the thesis of absolute freedom of speech through individual Self-Fulfillment are disregarded, social media, due to their operational structure, would not be suitable as the conducive environment for the Self-Fulfillment of individuals.

The operation of these platforms favors content that generates engagement through emotional reactions, to the detriment of information quality, reinforcing predispositions and isolating users from divergent perspectives. This dynamic subverts the notion of an open and democratic public space, limiting the potential of social media to promote inclusive dialogue and genuine citizen participation, instead contributing to social fragmentation and the consolidation of diffuse narratives and fake news.

In conclusion, the intrinsic characteristics of social media present considerable obstacles to the Scanlonian conception of autonomy linked to freedom of expression. The feasibility of users exercising such freedom in an informed and deliberative manner is impaired by phenomena such as echo chambers, the spread of misinformation, and the absence of a public sphere in traditional terms. These elements significantly distort the process by which information is received and processed, compromising individuals' capacity for autonomous and informed decision-making in the digital environment.

Bibliographic Reference

Baker, C. E. (1989), *Human Liberty and Freedom of Speech*. Oxford: Oxford University Press.

Baker, C. E. (1982), "Realizing Self-Realization: Corporate Political Expenditures and Redish's 'The Value of Free Speech'", *University of Pennsylvania Law Review*, 130 (3), pp. 646-677.

Barendt, E. (2007), Freedom *of Speech*. 2nd. Oxford: Oxford University Press.

Bresolin, K. (2023a), Liberdade de expressão, esfera pública, plataformas digitais e mídias sociais, *Ethic@ - An international Journal for Moral Philosophy*, 22 (2), pp.761-790.

Bresolin, K. (2023b), Da busca da verdade ao discurso de ódio: Desconstruindo o mito da absolutidade da liberdade de expressão na era digital. *Revista Sapere Aude*, 14(28), pp.465-491.

Espinoza, D. S. E. (2017), "A doutrina do mínimo existencial", *Interfaces Científicas – Humanas e Sociais*, 6 (1), pp.101-112.

Gomes, I. (2023), "Pobreza cai para 31,6% da população em 2022, após alcançar 36,7% em 2021", *Agência IBGE Notícias,* Recuperado de: https://abrir.link/CnEfQ, Accessed on: 30.03.2024.

Habermas, J. (2001). *Strukturwandel der Öffentlichkeit*: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft. Frankfurt am Main: Suhrkamp.

Habermas, J. (2022), "Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit" en Habermas, J., *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Berlin: Suhrkamp, pp.9-68.

Kruse, L. M.; Norris, D. R.; Flinchum, J. R (2017), "Social Media as a Public Sphere? Politics on Social Media", *The Sociological Quarterly*, 58 (1), pp.1-23.

Loader, B. D., y Mercea, D. (2011), "Networking Democracy? Social Media Innovations and Participatory Politics." *Information, Communication and Society,* 14 (6), pp.757-769.

Meiklejohn, A. (1948), *Free Speech and its Relation to Self-Government*. New York: Harper & Brother, 1948. Mill, J. S. (2015), "On Liberty" en Mill, J. S. *On Liberty, Utilitarianism and Other Essays*. Oxford: Oxford University Press, p.5-112.

Nagel, T. (1995), "Personal Rights and Public Space". *Philosophy Public Affairs,* 24 (2), pp.83-107.

Pariser, E. (2011), *The Filter Bubble*: What the Internet Is Hiding from You. New York: Penguin Press.

Pinotti, F. (2023), "Google retira mensagem contra a PL das Fake News da página inicial", *CNN Brasil,* Recuperado de: https://www.cnnbrasil.com.br/politica/google-retira-mensagem-contra-pl-das-fake-news-da-pagina-inicial/ . Accessed on: 02.04.2024.

Redish, M. "The Value of Free Speech", *University of Pennsylvania Law Review*, 130 (3), pp.591-645.

Samaržija, H. (2023), "The Epistemology of Fanaticism: Echo Chambers and Fanaticism" en Townsend, L., Tietjen, R. R., Schmid, H. B., & Staudigl, M. (eds.), *The Philosophy of Fanaticism: Epistemic, Affective, and Political Dimensions*. p. 69-87.

Scanlon, T. (1972). "A Theory of Freedom of Expression", Philosophy *& Public Affairs*", 1 (2), pp.204-226.

Scanlon, T. (2003), "Freedom of expression and categories of expression" en Scanlon, T., *The Difficulty of Tolerance*. Essay in Political Philosophy. Cambridge: Cambridge University Press, pp.84-112.

Shewale, R. (2024), "Social Media Users 2024 (Global Data & Statistics)", *Demandsage*. Recuperado de: https://www.demandsage.com/social-media-users/, Accessed on: 02.04.2024.

Shirky, C. (2011), "The Political Power of Social Media Technology, the Public Sphere, and Political Change", *Foreign Affairs*, 90 (1), pp.1-9. Recuperado de: https://faculty.cc.gatech.edu/~beki/cs4001/Shirky.pdf.

Sunstein, C. R. (2006), *Infotopia*: How Many Minds Produce Knowledge: New York: Oxford University Press.

Sunstein, C. R. (2017), *#Republic*: divided democracy in the age of social media. Princeton: Princeton University Press.

OPORTUNIDADES Y RIESGOS DE LOS NUEVOS CONTEXTOS DIGITALES

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 73-89
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.609851

Microtargeting político y vigilancia social masiva: impactos negativos en las democracias occidentales*

Political microtargeting and mass social surveillance: negative impacts on western democracies

*CARLOS SAURA GARCÍA***

Resumen: Este artículo se centra en los peligros para los procesos democráticos de uno de los principales instrumentos de propaganda política, el llamado *microtargeting* político. Este tipo de *microtargeting* permite dirigir contenidos específicos, hacia votantes específicos, en momentos específicos y vincularlos directamente con sus características, sesgos y vulnerabilidades individuales. El objetivo de este artículo es exponer el funcionamiento y los diversos tipos de *microtargeting* político y mostrar las posibles consecuencias nocivas de esta técnica en los procesos democráticos. Para lograr este objetivo, por una parte, se detallará la extracción, explotación y utilización de grandes conjuntos de datos para la creación de diversos tipos de propaganda política personalizada y, por otra parte, se analizarán las diversas propuestas existentes para limitar los

Abstract: This article focuses on the dangers to democratic processes of one of the main instruments of political propaganda, the so-called political microtargeting. This type of microtargeting allows you to direct specific content, towards specific people, at specific times and link them directly to their individual characteristics, biases, and vulnerabilities. The objective of this article is to expose the operation and the various types of political microtargeting, show the harmful consequences of this technique on democratic processes and propose solutions to address these negative consequences. To achieve this objective, on the one hand, the extraction, exploitation, and use of large data sets for the creation of various types of personalized political propaganda will be detailed and, on the other hand, the various existing proposals will be analyzed to limit the

Recibido: 26/03/2024. Aceptado: 18/06/2024.

* Esta publicación es parte del proyecto PID2022-139000OB-C22, financiado por MCIU/AEI/10.13039/501100011033/FEDER, UE y ha sido posible gracias a la financiación recibida de la Universitat Jaume I a través de un contrato predoctoral (PREDOC/2022/08).

** Investigador predoctoral en el departamento de Filosofía y Sociología de la Universitat Jaume I (UJI) de Castellón. Mis líneas de investigación se centran en el estudio de los efectos de las nuevas tecnologías del *big data* sobre la sociedad, profundizando especialmente en tres campos: retos éticos de la *datificación* y la hiperconectividad, implicaciones de los GAMAM (Google, Amazon, Meta, Apple y Microsoft) en el tratamiento y utilización de los datos masivos e implicaciones éticas de la utilización de los datos masivos en las contiendas democráticas y en los sistemas democráticos (*mass democracy*). Publicaciones recientes: Saura García, C. (2023). El big data en los procesos políticos: hacia una democracia de la vigilancia. *Revista de filosofía*, *80*, 215-232. https://doi.org/10.4067/S0718-43602023000100215 y Saura García, C. (2024). Digital expansionism and big tech companies: consequences in democracies of the European Union. *Humanities and Social Sciences Communications*, *11*(448), 1-8. https://doi.org/10.1057/s41599-024-02924-7

efectos negativos que puede causar el *microtargeting* político y para mejorar el funcionamiento de las democráticas occidentales.
Palabras clave: microtargeting, ecosistemas ciberfísicos, dataficación, capitalismo de la vigilancia, vigilancia social masiva

negative effects that can cause political microtargeting and to enhance the functioning of Western democracies.
Keywords: microtargeting, cyberphysical ecosystems, datafication, surveillance capitalism, mass social surveillance

Introducción

Los llamados ecosistemas *ciberfísicos* se han convertido en los elementos clave de la infraestructura digital de las sociedades modernas. Estos hacen posible la hibridación de los fenómenos de la digitalización, la *hiperconectividad*, la *dataficación* y la *algoritmización* y permiten la obtención de valor a partir de grandes conjuntos de datos (Calvo, 2021). Estos ecosistemas se podrían definir como la recreación de espacios artificiales solapados a aplicaciones, programas, dispositivos e instrumentos informáticos, biométricos o sensoriales interconectados y controlados por algoritmos que hacen posible la *hiperconectividad* y la *dataficación* de la totalidad de la realidad física y digital del comportamiento, las actividades, y las preferencias de personas, animales, objetos o procesos (Armenteras et al., 2016; Calvo, 2020). Los procesos de estos ecosistemas *ciberfísicos* permiten generar de forma continua grandes cantidades de datos; almacenar y procesar los conjuntos de datos para convertirlos en información, conocimiento y valor; y finalmente por medio de algoritmos aplicar esta información, conocimiento y valor para afectar e incidir en la realidad (Calvo, 2021).

Las grandes corporaciones tecnológicas del planeta (Google, Amazon, Meta, Apple, Microsoft, Alibaba, Baidu, Tencent, Huawei, etc.) obtienen grandes cantidades de beneficios económicos, vinculados en la mayoría de casos a la publicidad, gracias a la implementación del modelo de negocio del capitalismo de la vigilancia y a la explotación comercial y publicitaria de los ecosistemas *ciberfísicos* (Zuboff, 2020). Se estima que el valor del mercado de la publicidad global durante el año 2019 fue de 319 billones de dólares y se espera que este valor alcance los 1089 billones de dólares en 2027 (Galli et al., 2022). Pero esta explotación no está exenta de aspectos dañinos para la sociedad y especialmente para la democracia. Deibert (2019) subraya que este modelo de negocio esconde "tres dolorosas verdades". La primera es que este mercado está construido en base a la invasión de la privacidad de los ciudadanos y a su monetización. La segunda es que los propios ciudadanos son cómplices y conscientes de este nivel de vigilancia social masiva realizada por estas corporaciones y de la explotación de sus datos. Y la tercera y última, es que el modelo de negocio publicitario no es ni mucho menos incompatible con la manipulación y el autoritarismo. En relación a los efectos dañinos para la democracia, Calvo (2019) expone que:

> El neuromarketing político y económico, por ejemplo, puede utilizar los ecosistemas *ciberfísicos* para diseñar e implementar campañas altamente adictivas, capaces de modular y/o manipular la voluntad libre de los sujetos del sistema, así como de menguar o inhibir la capacidad crítica de votantes, gobernantes y clientes con el objetivo de maximizar el beneficio particular de unos pocos. (p.12)

Este fragmento pone en evidencia el enorme potencial manipulativo que pueden atesorar los ecosistemas *ciberfísicos*. Los diferentes actores interesados en influir en los procesos electorales y en el sistema democrático en general se han dado cuenta del enorme potencial de los grandes conjuntos de datos y han desarrollado innovadoras tecnologías para adaptar sus mensajes propagandísticos (Hersh, 2015; Wylie, 2019; Da Empoli, 2020; Schick, 2020; Woolley, 2023). Esta situación ha provocado la creación de una infraestructura de propaganda computacional capaz de crear contenidos altamente personalizados para influir específicamente en determinados segmentos de la sociedad basándose en los patrones extraídos de los grandes conjuntos de datos de los ecosistemas *ciberfísicos* (Howard, 2020; Dawson, 2021; Woolley, 2023).

La rápida evolución de los ecosistemas *ciberfísicos* ha hecho posible la aparición del denominado *microtargeting* político en estos entornos (Woolley y Howard, 2018). El *microtargeting* político es un tipo de propaganda computacional automatizada y *algoritmizada* que tiene el objetivo de seleccionar a ciudadanos específicos y ofrecerles una propaganda política activa que pueda influir y manipular su comportamiento, sus opiniones y su ideología. La principal característica de esta propaganda es su capacidad de aprender a partir de las interacciones con los ciudadanos y afinarse a través del dialogo virtual entre los propios ciudadanos y los contenidos de la propaganda computacional (Bashyakaria et al., 2019)[1].

El objetivo de este artículo será exponer los efectos negativos que el *microtargeting* político puede tener sobre los procesos democráticos de las sociedades modernas. En un primero momento, se analizará los procesos de extracción y explotación de datos que alimentan la tecnología del *microtargeting*. A continuación, se expondrán los diversos tipos de *microtargeting* y las peculiaridades de cada uno de ellos. Finalmente se profundizará en la situación existente en las democracias de las sociedades occidentales y se analizaran las diversas propuestas existentes para limitar los efectos negativos causados por el *microtargeting* político y mejorar el funcionamiento de los sistemas democráticos.

1. Vigilancia social masiva: extracción y explotación de grandes conjuntos de datos

Actualmente el *microtargeting* está estrechamente vinculado con los fenómenos de la vigilancia social masiva y la recolección indiscriminada de datos (Dawson, 2021, 2023). El modelo de negocio del capitalismo de la vigilancia ha desarrollado un tejido de vigilancia social masiva centrado en la *datificación* de la totalidad de los aspectos de la ciudadanía y de las sociedades con el objetivo de capturar la mayor cantidad de datos posibles de todas las acciones que se realizan en el mundo para posteriormente extraer valor de estos y obtener beneficios económicos (Ebeling, 2022).

La aplicación práctica de este tejido de vigilancia social masiva supone la extracción de datos de una vasta cantidad de aspectos, actividades, comportamientos y procesos del día a día de la ciudadanía, como por ejemplo la *datificación* de la información relativa a navega-

1 Es importante destacar que el embrión del *microtargeting* político se encuentra en la disciplina de la psicotecnia. Schumpeter (1958) fue uno de los primeros pensadores en exponer que la psicotecnia electoral —basada en técnicas y métodos psicológicos y sociológicos— podía ser utilizada para influir y moldear el comportamiento electoral en procesos democráticos.

ción en la red (Bashyakaria et al., 2019), las compras online y en la vida real (Llaneza, 2019) o las operaciones realizadas con tarjetas de crédito (Mayer-Schönberger y Cukier, 2013). Pero también otros procesos extractivos más dañinos para la privacidad y la intimidad de la ciudadanía como son la *datificación* de la intimidad corporal (Farahany, 2023), la voz (Turow, 2021), los rostros (De Miguel, 2024), la ubicación de los smartphones (Thompson y Warzel, 2019) y la extracción de datos y metadatos privados a través de aplicaciones destinadas al entretenimiento de niños (Fowler, 2022a) o a través de diversos dispositivos digitales como son el asistente virtual, la Smart TV, la nevera, el microondas, la aspiradora, las luces inteligentes e incluso el inodoro (Fowler, 2022b).

La gran cantidad de datos y metadatos extraídos de forma continua por parte de la maquinaria del capitalismo de la vigilancia ha otorgado la posibilidad a cualquier persona con unos conocimientos mínimos de navegación en la red y de análisis de datos de tener "direct access to the minds and lives of guards, clerks, girlfriends… a detailed trail of personal information that would perviously have taken months of careful oberservation to gather" (Wylie, 2019: 49). Un claro ejemplo de la facilidad de explotar, segmentar y extraer valor de grandes conjuntos de datos es el trabajo final de grado que realizaron dos estudiantes de la Escuela de Ingeniería y Ciencias Aplicadas John A. Paulson de Harvard (Zewe, 2020). Estos estudiantes fueron capaces a través de un conjunto de datos robados a la compañía de informes de crédito Experian con información privada de 6 millones de personas en relación con 69 variables diferentes (dirección, número de teléfono, donaciones a partidos políticos, correo electrónico, contraseñas, hijos, etc.) que encontraron en la *dark* web de descubrir y recrear perfiles característicos de tres senadores, tres diputados de la Cámara de Representantes y el alcalde de Washington DC y sus miembros de gabinete, en los cuales se incluyan informaciones como las calificaciones crediticias, los números de teléfono y las direcciones. La facilidad con la que estos estudiantes totalmente amateurs fueron capaces de identificar características de individuos específicos pone en evidencia el potencial que pueden tener la realización de este tipo de técnicas por profesionales y corporaciones con una tecnología mucho más potente y con unos conjuntos de datos mucho más grandes y actualizados.

En el campo de la política este tipo de técnicas permite atomizar e *hipersegmentar* grandes bases de datos e identificar a personas específicas en contextos específicos para aplicarles propaganda política personalizada, es decir, *microtargeting* político, con el objetivo de persuadirlas y manipularlas políticamente de la forma más precisa y eficaz posible (Matz et al., 2017; Nave et al., 2018; Kosinski, 2021).

Las primeras referencias de utilización de *microtargeting* político en entornos *ciberfísicos* datan de las campañas electorales presidenciales de Barack Obama en 2008 y 2012[2]. En estas dos contiendas se construyó y se desarrolló una infraestructura de creación de perfiles en base a las características de las personas y de búsqueda de perfiles específicos a partir

2 Cabe señalar que el *microtargeting* político ya existía previamente a la utilización masiva de las plataformas digitales por medio de técnicas de marketing directo. Las primeras referencias de utilización de marketing directo datan de las décadas de 1960 y 1970 en los EEUU. Estas técnicas fueron introducidas por el publicista Richard Viguiere en el Partido Republicano y se enfocaron a la creación de bases de datos de direcciones de los ciudadanos, la segmentación de votantes y la personalización de mensajes (Moriyama, 2022). Sus efectos fueron determinantes para la victoria del candidato republicano Ronald Reagan en las elecciones presidenciales de 1980.

de la adquisición de conjuntos de datos a las grandes corporaciones tecnológicas y a *data brokers* (Issenberg, 2012; Bimber, 2014; Gerodimos y Justinussen, 2015). Esta infraestructura era aún embrionaria y bastante básica, pero fue ampliamente desarrollada y utilizada por el equipo de campaña de Donald Trump en las elecciones presidenciales de 2016 y por el equipo del Brexit en el referéndum de permanencia de Reino Unido en la Unión Europea (UE) en 2016 (Kaiser, 2019; Wylie, 2019).

En estas dos contiendas la empresa especializada en operaciones de comunicación estratégica Cambridge Analytica fue capaz de recrear los perfiles psicológicos y las afinidades políticas de millones de personas actualizadas al instante gracias a sus perfiles de Facebook (Kaiser, 2019; Wylie, 2019)[3]. El *whistleblower* de la empresa Cambridge Analytica, Christopher Wylie, reveló que estos perfiles psicológicos y afinidades políticas fueron utilizados para realizar multitud de campañas de *microtargeting* político para persuadir y manipular a la población por medio de las denominadas operaciones psicológicas (Wylie, 2019)[4]. El estrecho margen de la victoria de Donald Trump en algunos importantes estados (diferencias menores al 1% de los votos en los estados de Michigan, Wisconsin y Pensilvania), y de la victoria del Brexit (diferencia de un 3,78% entre los votos a favor y en contra), ponen de manifiesto lo decisivas que pudieron ser estas campañas de *microtargeting* para el resultado final de estas contiendas (Jamieson, 2018; Wylie, 2019).

Durante los últimos 5 años algunas grandes corporaciones digitales y algunos estados han empezado a desarrollar las reglamentaciones en relación a privacidad y protección de los datos de sus usuarios y sus ciudadanos (Galli, 2021; Fässler, 2023), especialmente después del descubrimiento por parte de la opinión pública en 2018 de las prácticas de extracción de datos realizada de forma ilícita por Cambridge Analytica a través de Facebook y la puesta en marcha de propaganda computacional basada en psicometría (Cadwalladr, 2017; Cadwaladr y Graham-Harrison, 2018; Rosenberg et al., 2018). Como consecuencia de la actualización y desarrollo de las reglamentaciones puestas en marcha por las grandes corporaciones digitales y los gobiernos, se han buscado nuevas formas de extraer datos y metadatos de la ciudadanía para posteriormente poner en marcha campañas de persuasión y manipulación política a través de *microtargeting*.

En el caso de las elecciones presidenciales de los Estados Unidos (EEUU) de 2020, tanto Joe Biden como Donald Trump utilizaron *apps* oficiales de campaña con la finalidad de extraer datos de los ciudadanos (Woolley y Gursky, 2020). En el ciclo político americano iniciado en 2016 los candidatos utilizaron las grandes plataformas digitales para extraer datos y manipular a los votantes (Kaiser, 2019; Wylie, 2019). En el ciclo político que empezó en 2020 esta función se ha trasladado a las aplicaciones oficiales de campaña. Woolley y Gursky (2020) subrayan que estas *apps*:

3 Esto fue posible gracias a una filtración de datos privados de 87 millones de usuarios de Facebook en 2014 y a la combinación de estos datos con bases de datos de consumidores adquiridas a compañías especializadas en la compra-venta de datos (Kaiser, 2019; Wylie, 2019).

4 Durante este tiempo la empresa Cambridge Analytica también realizó campañas de manipulación electoral en otros países como Nigeria, Trinidad y Tobago, Moldavia o Ucrania (Wylie, 2019).

> […] allow the Trump and Biden teams to speak directly to likely voters. They also allow them to collect massive amounts of user data without needing to rely on major social-media platforms or expose themselves to fact-checker oversight of particularly divisive or deceptive messaging.

Este fragmento pone en evidencia que el principal objetivo de estas aplicaciones es la extracción indiscriminada de datos y metadatos de sus usuarios, y de los contactos de estos usuarios, sin depender de terceras partes. Los permisos de extracción de datos solicitados por estas plataformas para poder ser utilizadas, en el caso de la *app* de Joe Biden, incluían obligatoriamente el número de teléfono, el código postal, el correo electrónico, los contactos, la información de las búsquedas en internet o las redes Wifi; mientras que en el caso de la *app* de Donald Trump a esta lista se le sumaban multitud de permisos de extracción de datos como son la ubicación, la obtención de información privada del propio *smartphone* y de la tarjeta SD o la información relativa al Bluetooth (Woolley y Gursky, 2020).

Las aplicaciones de campaña se han convertido en un elemento más de la infraestructura del capitalismo de la vigilancia dedicada a afectar a los procesos democráticos. Estas aplicaciones permiten obtener a los candidatos flujos de datos actualizados directamente de sus simpatizantes más afines y de todos sus contactos, y posteriormente combinarlos con multitud de conjuntos de datos comprados a las grandes corporaciones digitales o a empresas dedicadas al negocio de compra-venta de datos como Acxiom, Resonate, i360 o Targetsmart para de esta forma crear microsegmentos de ciudadanos con características similares e identificar a personas específicas en contextos específicos para intentar persuadirlas y manipularlas con *microtargeting* político (Bartlett et al., 2018; Bashyakaria et al., 2019; Woolley y Gursky, 2020).

2. La sala de máquinas del *microtargeting* político

A partir del proceso continuo de extracción y explotación de grandes conjuntos de datos de la ciudadanía y de la sociedad en general, los algoritmos proceden a crear, seleccionar y enviar las informaciones más adecuados para persuadir y manipular a cada persona. En este punto es fundamental enviar de la forma más sutil, eficiente y eficaz posible argumentos, contenidos y propaganda política que se adapten al máximo posible a los intereses personales, sociales, psicológicos o emocionales de cada persona, a los diversos contextos y circunstancias de la vida diaria de la ciudadanía y a las variaciones de estos. El objetivo del *microtargeting* político es persuadir y manipular la opinión, la ideología y la intención de voto de determinados ciudadanos en favor de los intereses de quien lo ejecuta (Bashyakaria et al., 2019). Este tipo de *microtargeting* abarca diversos métodos para lograr, por una parte, unos niveles de adaptación y personalización máximos y, por otra parte, unos niveles de persuasión y manipulación máximos para cada persona (Matz et al., 2017; Bashyakaria et al., 2019; Bakir, 2020; Iyer et al., 2021). Entre estos métodos destacan:

— El *A/B testing*
— El *geotargeting*
— El *psicotargeting*
— El targeting cognitivo
— El *neurotargeting*

El *A/B testing* hace referencia a un tipo de *microtargeting* en el cual se comparan dos o más variables de un mismo argumento, contenido o propaganda para, de esta forma, maximizar el impacto que tiene sobre cada ciudadano (Bashyakaria et al., 2019)[5]. El *A/B testing* utiliza los datos y metadatos extraídos de la ciudadanía y la sociedad para, en un primer momento, seleccionar a determinados segmentos de la ciudadanía a los cuales va a afectar y, posteriormente, optimizar los contenidos para incrementar su poder de persuasión y manipulación (Bashyakaria et al., 2019). El objetivo de este método es interactuar con los votantes y adaptarse a sus características a través de contenidos automatizados únicos, personalizados y actualizados constantemente gestionados por algoritmos, los cuales crean, combinan y modifican una amplia variedad de botones, textos, imágenes y videos para aumentar la eficacia y la eficiencia de los contenidos enviados.

El *geotargeting* está vinculado con la utilización de datos y metadatos acerca de la ubicación para enviar argumentos y contenidos determinados y propaganda personalizada (Iyer et al., 2021). La combinación de los datos de la ubicación de las personas con otros conjuntos de datos, como pueden ser datos en relación con actividades físicas, compras, características sociales, contactos o actividad en las redes sociales, permite conocer los intereses, opiniones, costumbres y deseos de cada persona en cada momento e incrementar la precisión de los argumentos, contenidos y propaganda enviados. Las dos principales variantes de aplicación del *geotargeting* son el *geofencing* y el *IP targeting*.

El *geofencing* fija un perímetro virtual alrededor de un determinado punto y promociona un determinado mensaje con unas características específicas a las personas que se encuentran en el interior o que se acercan hacia él gracias a sus datos en relación a ubicación, Bluetooth, etc. (Bashyakaria et al., 2019; Woolley y Gursky, 2020; Iyer et al., 2021). El *IP targeting* se basa en el envío de mensajes con argumentos y contenidos personalizados a direcciones IP de determinados dispositivos conectados a la red especialmente seleccionadas para lograr unos niveles de persuasión y manipulación máximos en las personas que los usan (Bashyakaria et al., 2019; Iyer et al., 2021; Singer, 2022).

El *psicotargeting* es el tipo de *microtargeting* que utiliza la información psicológica de la ciudadanía para crear perfiles psicológicos de cada persona, buscar perfiles específicos entre la población y personalizar, optimizar y mejorar las campañas de persuasión y manipulación política basándose en las peculiaridades psicológicas de cada ciudadano (Matz et al., 2017; Bashyakaria et al., 2019; Bakir, 2020). La posibilidad de analizar grandes conjuntos de datos heterogéneos de cada ciudadano gracias a innovadoras herramientas psicométricas ha permitido predecir de forma detallada y actualizada la personalidad, los estados de ánimo o las emociones de cada persona en cada momento[6]. Entre estas herramientas destacan la predicción de la personalidad a través de *likes* de Facebook (Kosinski et al., 2013; Youyou et al., 2015), las preferencias musicales (Nave et al., 2018) o una imagen del rostro de las

5 Un ejemplo de utilización de *A/B testing* es la utilización por parte de la candidatura de Donald Trump a las elecciones presidencial de EEUU de 2016 de entre 50.000 y 60.000 variaciones diarias de la mismo anuncio político en Facebook (Bartlett et al., 2018; Da Empoli, 2020).

6 Estas herramientas psicométricas en la mayoría de los casos toman como referencia el modelo OCEAN, también llamado modelo de los cinco grandes rasgos de personalidad, en el cual se valora la personalidad de cada persona teniendo en cuenta la apertura, responsabilidad, extroversión, amabilidad e inestabilidad emocional (Gerber et al., 2011).

personas (Kosinski, 2021; De Miguel, 2024). En muchos casos el *psicotargeting* político se lleva a la práctica por medio de las llamadas operaciones psicológicas (*psychological operations* o PSYOPS en inglés). Las PSYOPS son un tipo de propaganda, normalmente utilizada por ejércitos en conflictos bélicos, de tipo manipulativo que incide en las vulnerabilidades emocionales y en los estados de ánimo (Haig y Hajdu, 2017).

El targeting cognitivo utiliza los datos vinculados con los sesgos cognitivos, es decir, los errores imperceptibles del cerebro a la hora de generar interpretaciones informativas incompletas, imperfectas y defectuosas que dan lugar a formas de actuar y opinar irracionales, para enviar argumentos, contenidos y propagandas determinadas con el objetivo de afectar de forma disimulada a la ciudadanía e influir y manipular de forma decisiva su ideología y sus opiniones (Kahneman, 2011; Mercier y Sperber, 2017; Sanborn y Harris, 2018). Entre los sesgos cognitivos más comunes destacan el sesgo de punto ciego, el sesgo de apoyo a la elección, el sesgo de anclaje, el sesgo de confirmación, la ilusión de agrupamiento, el efecto *Bandwagon*, el sesgo innovativo o el sesgo de actualidad (Kahneman, 2011; Juárez Ramos, 2019).

El *neurotargeting* está basado en la extracción y explotación de los datos de la intimidad corporal de las personas, es decir, datos vinculados a la información biométrica e íntima como son las ondas cerebrales, la presión sanguínea, la actividad electrodérmica o la electromiografía facial[7]. En el ámbito de la política esta técnica busca crear marcos, contenidos, estímulos o impulsos con el objetivo de persuadir y manipular a las personas e influir en la autonomía, la cognición y la libertad de los ciudadanos de forma prácticamente imperceptible.

La puesta en marcha de un *microtargeting* político capaz de crear multitud de variables de un mismo contenido, de actualizar y optimizar estas variables dependiendo de la ubicación de las personas, de las peculiaridades psicológicas, de beneficiarse de las limitaciones cognitivos y de utilizar información biométrica e íntima ha permitido crear una propaganda computacional automatizada y *algoritmizada* basada en la personalización, la imperceptibilidad, la evasión de la cognición y la potenciación del pensamiento irracional capaz de manipular a la ciudadanía y poner en serio riesgo los principios básicos de la democracia (Han, 2021). La utilización del *microtargeting* político puede manipular la voluntad política de la ciudadanía, afectar directamente a la soberanía de la ciudadanía y adulterar los procesos democráticos (Da Empoli, 2020; Zuboff, 2020; Han, 2021). Ante esta amenaza para la democracia, algunos estados han empezado a fortalecer y desarrollar la reglamentación en relación con la extracción y análisis de grandes conjuntos de datos y con el envío de propaganda computacional personalizada (Galli et al., 2022; Fässler, 2023).

---

7 A pesar de que en este trabajo el *neurotargeting* se considera un método más de *microtargeting*, existe una diferencia notable entre el resto de los métodos de *microtargeting* y el *neurotargeting*. El resto de los métodos de *microtargeting* se basan en la extracción y explotación de datos de la ciudadanía creados de forma consciente por esta, mientras que el *neurotargeting* se cimienta en la extracción y explotación de conjuntos de datos de la intimidad corporal que no han sido creados de forma consciente por las personas.

3. El *microtargeting* político en las democracias occidentales

La capacidad del *microtargeting* político de realizar una propaganda computacional automatizada, *algoritmizada*, personalizada, e incluso sintética, permite dirigir contenidos específicos, hacia votantes específicos, en momentos específicos y vincularlos directamente con sus características intimas, sus sesgos y sus vulnerabilidades para lograr una persuasión y manipulación política máxima (Bashyakaria et al., 2019). Además de estas características, que ya de por sí adulteran los procesos democráticos y ponen en serio riesgo los principios básicos de las democracias deliberativas occidentales (Habermas, 2021), estos efectos negativos se pueden agravar a través de la difusión de contenidos ligados con informaciones de dudosa veracidad, desinformaciones, *fake news o deep fakes* (D'Ancona, 2019; Howard, 2020; Nightingale y Farid, 2022; Woolley, 2023) o por medio del bombardeo de propaganda a determinados colectivos con rasgos antisociales, impulsivos o agresivos para que estos se inflamen y creen un clima de distorsión social "(Garcia y Sikström, 2014; González Moraga, 2015; Jamieson, 2018; Wylie, 2019). Dawson (2021) resume esta situación diciendo que el *microtargeting* político:

> allows individual-level messaging to be deployed to influence voting behavior and is able to be leveraged for more insidious dis/misinformation campaigns. What started as a way for businesses to connect directly with potential customers has transformed into a disinformation machine at a scale that autocratic governments of the past could only imagine. (p.64)

Este fragmento enfatiza los principales efectos negativos y la peligrosidad del *microtargeting* político para el sistema democrático. Estas cuestiones se acrecientan como consecuencia del desconocimiento de la ciudadanía de la extracción y explotación de sus datos para la realización de *microtargeting* político y de las características de este fenómeno (Llaneza, 2019). La mayoría de los ciudadanos no tienen en consideración que los contenidos políticos a los cuales están expuestos en los ecosistemas *ciberfísicos* no son objetivos y están fuertemente influenciados y sesgados en favor de los intereses del anunciante (Fässler, 2023).

La puesta en marcha de *microtargeting* político provoca impactos negativos en los pilares fundamentales de las democracias deliberativas occidentales principalmente a través de dos ámbitos. Por una parte, la monitorización, extracción y explotación de grandes conjuntos de datos de la privacidad e intimidad de cada persona socava la integridad, la dignidad, la personalidad y el anonimato de cada persona dando lugar a la posibilidad de monitorizar sus comportamientos y controlar sus acciones y actividades (Zuboff, 2020; Han, 2021; Varoufakis, 2023). Por otra parte, la capacidad de creación de contenidos totalmente personalizados de forma automatizada y *algoritmizada* para cada individuo —basados en la explotación de los datos privados e íntimos de las personas— erosiona el conocimiento y la veracidad necesarias para el correcto funcionamiento de democracia (Woolley, 2023). La combinación de estos factores hace posible que los individuos, empresas, organizaciones o instituciones que ponen en marcha el *microtargeting* político puedan intervenir y condicionar la soberanía, autonomía y autodeterminación de la ciudadanía y producir una falta de agencia epistémica

que ponga en cuestión el control de la ciudadanía sobre la construcción de su propia ideología y sus propias convicciones políticas (Coeckelbergh, 2022, 2024).

Para hacer frente a los efectos dañinos del *microtargeting* político en los procesos democráticos de las sociedades occidentales, actualmente existen dos enfoques reglamentaritos totalmente diferentes: el de la UE y el de los EEUU.

La UE tiene una clara voluntad de regular y limitar los efectos dañinos del *microtargeting* en los sistemas democráticos y ha desarrollado un amplio paquete de regulaciones para este propósito (Galli et al., 2022; Fässler, 2023). Entre estas regulaciones destacan el GDPR (General Data Protection Regulation), la DSA (Digital Services Act) y la AIA (Artificial Intelligence Act) ya que afrontan la extracción de datos, la transparencia, la segmentación y la personalización de los anuncios políticos.

En primer lugar, el GDPR ha hecho obligatoria para cualquier organización que extraiga y explote datos la obtención de un consentimiento «valido» de las personas afectadas a la hora de extraer y usar datos privados y también introduce la posibilidad de que los ciudadanos puedan optar por limitar la explotación de sus datos privados para la realización de propaganda directa. En segundo lugar, la DSA, por una parte, obliga a las grandes plataformas digitales a evaluar los riesgos de los contenidos de sus plataformas para los procesos electorales, la seguridad pública y la desinformación y a tomar medidas para mitigarlos y a aumentar la transparencia y control en relación con la propaganda política. Por otra parte, afronta directamente los problemas del *microtargeting* político prohibiendo las técnicas de segmentación y personalización que impliquen la utilización de datos personales específicos como puedan ser datos vinculadas con la salud, la raza o las creencias religiosas. Finalmente, la AIA, por una parte, busca que el uso de la inteligencia artificial generativa en el ámbito del *microtargeting* sea transparente, ética y responsable y, por otra parte, que esta no pueda ser utilizada para socavar los derechos y libertades fundamentales de los ciudadanos y los procesos democráticos.

El tejido regulativo de la UE es uno de los más desarrollados del mundo, por no decir el más desarrollado, en materia de protección de datos, transparencia, consentimiento y lucha contra la manipulación política. A pesar de esto, la ciudadanía europea aún está expuesta a determinados riesgos que pueden socavar los procesos democráticos. Por una parte, algunas de las grandes corporaciones digitales estadounidenses, como Google, Meta y Amazon, han sido multadas y amonestadas por incumplir las prerrogativas del GDPR y de la DSA y por transferir de forma continuada grandes conjuntos de datos a EEUU para evitar las reglamentaciones de la UE (Commission Nationale de l'Informatique et des Libertés, 2022; European Data Protection Board, 2022; Perez Colome y Ayuso, 2023). Por otra parte, en las políticas de privacidad de las grandes corporaciones digitales chinas se indica que los conjuntos de datos de los europeos pueden ser trasladados y explotados fueras de las fronteras de la UE sin tener en cuenta las leyes de la UE, además de poder ser utilizados por el propio gobierno chino (Hoffman y Attrill, 2021). Por último, la prohibición de las técnicas de segmentación y personalización que impliquen la utilización de datos personales específicos plasmadas en DSA pueden ser sorteadas por medio de la obtención del consentimiento —en la mayoría de casos a través de un modo no totalmente informado— de las personas afectadas (Fässler, 2023).

En EEUU no hay ninguna regulación estatal en relación con la extracción masiva de datos, la inteligencia artificial generativa, el *microtargeting* político y los efectos que el capitalismo de la vigilancia puede causar en los procesos democráticos (Dawson, 2021, 2023) y además se está impidiendo judicialmente al gobierno realizar cualquier tipo de regulación de este tipo (Zakrzewski, 2023). En relación con los grandes conjuntos de datos, en EEUU los datos privados pueden ser extraídos, comprados, transferidos o explotados por cualquier personaje nacional o extranjero sin prácticamente limitaciones. En relación con el *microtargeting* político, existen algunas limitaciones vinculadas con la realización de propaganda política personalizada por parte de la administración y del gobierno de los EEUU, pero ninguna limitación en las técnicas, los contenidos o los objetivos del resto de actores (candidatos, partidos políticos, empresarios, poderes fácticos, gobiernos extranjeros, etc.) (Dawson, 2023). Las únicas limitaciones existentes en este caso son las débiles e interesadas autorregulaciones introducidas por algunas de las grandes corporaciones digitales fuertemente dependientes del mercado de la publicidad (Fässler, 2023).

En el caso de la UE, a pesar de los grandes esfuerzos regulativos realizados, los procesos democráticos y la ciudadanía aún se pueden ver influenciados y manipulados por el *microtargeting* político como consecuencia de, principalmente, dos cuestiones. En un primer momento, la llamada "paradoja de la privacidad" (Acquisti y Grossklags, 2005). Los ciudadanos europeos son conscientes de los riesgos de la extracción y explotación de sus datos para fines comerciales y propagandísticos, pero dan su consentimiento, para de esta forma, poder continuar utilizando las plataformas y los programas tecnológicos (Lastra-Anadón y Rubio, 2020; Solove, 2021). En un segundo momento, la transferencia de estos grandes conjuntos de datos al extranjero para poder explotarlos obviando las reglamentaciones europeas supone de facto la posibilidad de *hipersegmentar* a la ciudadanía europea, personalizar contenidos y realizar propaganda computacional automatizada y *algoritmizada* (Saura García, 2024).

La UE debe seguir desarrollando y actualizando sus regulaciones para hacer frente a los impactos negativos de la vigilancia social masiva y el *microtargeting* político sobre los sistemas democráticos provocados por la recopilación e intercambio de información y por la explotación de los sesgos y las vulnerabilidades de los ciudadanos. Para lograr este propósito, Galli et al. (2022) argumentan que se deben ampliar y desarrollar medidas específicas destinadas a garantizar el libre consentimiento en relación a la extracción y explotación de datos, limitar la posibilidad de otorgar el consentimiento para la extracción y explotación de datos a cambio de servicios y contraprestaciones y promover un ecosistema de intercambio de datos libre y justo que proteja los datos de los ciudadanos europeos.

En el caso de los EEUU, Dawson (2021) resume la situación diciendo que el ecosistema totalmente desregulado del capitalismo de la vigilancia y del *microtargeting* político estadunidense pone en serio riesgo la capacidad de opinar y reflexionar de la ciudadanía, la soberanía popular y el correcto funcionamiento de la democracia en nombre del libre mercado económico. De la misma forma, también aumenta desmesuradamente el poder económico y político de terceros actores, como son poderes facticos, gobiernos extranjeros, personajes multimillonarios, etc. (Da Empoli, 2020; Zuboff, 2020).

4. Conclusiones

La vigilancia, control y manipulación política de la ciudadanía a través de la extracción y explotación de sus datos y metadatos personales y de campañas de manipulación política automatizadas y *algoritmizadas*, entre las que destaca el *microtargeting,* pero también las cámaras de resonancia, los filtros de burbuja o la utilización masiva de *bots*, crean una opinión pública artificial y sintética y ponen en serio riesgo el correcto funcionamiento de los procesos y los sistemas democráticos en general (García-Marzá y Calvo, 2022, 2024).

El desarrollo exponencial de las técnicas, por una parte, de *datificación* y extracción de datos y metadatos y, por otra parte, de manipulación política en diferentes frentes, como son los *deep fakes*, la tecnología *deep learning* y el uso de inteligencia artificial generativa a un ritmo mucho más rápido que el de cualquier posible actualización de la regulación (Helmus, 2022; Millière, 2022), el incumplimiento de la reglamentación existente por parte de algunas de las corporaciones digitales más importantes del mundo (Saura García, 2024) y la indiferencia radical del modelo de negocio del capitalismo de la vigilancia por la libertad de expresión y los pilares de los sistemas democráticos (Zuboff, 2020) están transformando el actual formato democrático en una democracia de la vigilancia de carácter artificial, sintético, instrumental, sesgado y despótico.

La democracia de la vigilancia hace referencia a un formato democrático emergente en el que los procesos implicados quedan desvirtuados y vaciados de legitimidad como consecuencia del contexto de vigilancia, control y manipulación política que limita la capacidad de opinar y reflexionar de la ciudadanía por medio de la personalización, artificialización y sintetificación de la información que consume la ciudadanía y la destrucción de la opinión pública crítica y madura (Saura García, 2023; Calvo y Saura García, en prensa). El funcionamiento de la democracia de la vigilancia mantiene todo el poder de la democracia en la ciudadanía, pero de forma instrumental, puesto que la extracción de datos y la propaganda computacional política provocan que estos sean unas marionetas de los intereses políticos de las grandes corporaciones digitales, los poderes fácticos, los personajes multimillonarios o los gobiernos extranjeros.

Bibliografía

Acquisti, A., y Grossklags, J. (2005). Privacy and Rationality in Individual Decision Making. *IEEE Security and Privacy*, *3*(1), 26-33. https://doi.org/10.1109/MSP.2005.22

Armenteras, D., González, T. M., Vergara, L. K., Luque, F. J., Rodríguez, N., y Bonilla, M. A. (2016). A review of the ecosystem concept as a "unit of nature" 80 years after its formulation. *Ecosistemas*, *25*(1), 83-89. https://doi.org/10.7818/ECOS.2016.25-1.12

Bakir, V. (2020). Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Communication*, *5*(67), 1-16. https://doi.org/10.3389/fcomm.2020.00067

Bartlett, J., Smith, J., y Acton, R. (2018). *The Future of Political Campaigning*. Demos. Recuperado de https://ico.org.uk/media/2259365/the-future-of-political-campaigning.pdf

Bashyakaria, V., Hankey, S., Macintyre, A., Renno, R., y Wright, G. (2019). *Personal Data: Political Persuasion Inside the Influence Industry. How it works.* Tactical Tech's Data

and Politics. Recuperado de https://cdn.ttc.io/s/tacticaltech.org/Personal-Data-Political-Persuasion-How-it-works.pdf

Bimber, B. (2014). Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment. *Journal of Information Technology & Politics*, *11*(2), 130-150. https://doi.org/10.1080/19331681.2014.895691

Cadwaladr, C., y Graham-Harrison, E. (17 de marzo de 2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Recuperado de https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Cadwalladr, C. (14 de mayo de 2017). Follow the data: does a legal document link Brexit campaigns to US billionaire?. *The Guardian*. Recuperado de https://www.theguardian.com/technology/2017/may/14/robert-mercer-cambridge-analytica-leave-eu-referendum-brexit-campaigns

Calvo, P. (2019). Democracia algorítmica: consideraciones éticas sobre la dataficación de la esfera pública. *Revista del Clad. Reforma y Democracia*, *74*, 5-30.

Calvo, P. (2020). Democracia aumentada. Un ecosistema ciberético para una participación política basada en algoritmos. *Ápeiron. Estudios de Filosofía*, *12*, 129-141.

Calvo, P. (2021). El gobierno ético de los datos masivos. *Dilemata*, *34*, 31-49.

Calvo, P. y Saura García C. (en prensa). Democracia de la vigilancia: datos, activismo y contrapoder. *Revista Internacional de Pensamiento Político*, 19.

Coeckelbergh, M. (2022). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI and Ethics*, *3*(4), 1341-1350. https://doi.org/10.1007/S43681-022-00239-4

Coeckelbergh, M. (2024). *Why AI Undermines Democracy and What To Do About It*. Cambridge: Polity Press.

Commission Nationale de l'Informatique et des Libertés. (2022). Use of Google Analytics and data transfers to the United States: the CNIL orders a website manager/operator to comply. Recuperado de https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply

D'Ancona, M. (2019). *Posverdad: La nueva guerra en torno a la verdad y cómo combatirla*. Madrid: Alianza Editorial.

Da Empoli, G. (2020). *Los ingenieros del caos*. Madrid: Ediciones Anaya.

Dawson, J. (2021). Microtargeting as Information Warfare. *The Cyber Defense Review*, *6*(1), 63-80. https://doi.org/10.2307/26994113

Dawson, J. (2023). Who Controls the Code, Controls the System: Algorithmically Amplified Bullshit, Social Inequality, and the Ubiquitous Surveillance of Everyday Life. *Sociological Forum*, 1-24. https://doi.org/10.1111/SOCF.12907

De Miguel, R. (17 de junio de 2024). Cámaras con IA en el metro de Londres captan el estado emocional de los viajeros. *El País*. Recuperado de https://elpais.com/ciencia/2024-06-17/camaras-con-ia-en-el-metro-de-londres-captan-el-estado-emocional-de-los-viajeros.html

Deibert, R. J. (2019). The Road to Digital Unfreedom: Three Painful Truths About Social Media. *Journal of Democracy*, *30*(1), 25-39. https://doi.org/10.1353/JOD.2019.0002

Ebeling, M. F. E. (2022). *Afterlives of data : life and debt under capitalist surveillance*. Berkeley: University of California Press.

European Data Protection Board. (2022). Italian SA bans use of Google Analytics: no adequate safeguards for data transfers from Caffeina Media S.r.l. to the U.S. Recuperado de https://edpb.europa.eu/news/national-news/2022/italian-sa-bans-use-google-analytics-no-adequate-safeguards-data-transfers_en

Farahany, N. A. (2023). *The Battle for Your Brain*. New York: St. Martin's Press.

Fässler, M. (2023). *Google's Privacy Sandbox Initiative: Old wine in new skins* (N.º 2023/01). Recuperado de https://www.zora.uzh.ch/id/eprint/232978/1/Google_s_Privacy_Sandbox.pdf

Fowler, G. A. (6 de septiembre de 2022a). Your kids' apps are spying on them. *The Washington Post*. Recuperado de https://www.washingtonpost.com/technology/2022/06/09/apps-kids-privacy/

Fowler, G. A. (12 de octubre de 2022b). Tour Amazon's dream home, where every appliance is also a spy. *The Washington Post*. Recuperado de https://www.washingtonpost.com/technology/interactive/2022/amazon-smart-home/

Galli, F. (2021). *Algorithmic business and EU law on fair trading*. Università di Bologna. Recuperado de http://amsdottorato.unibo.it/9750/1/tesifinale_galli.pdf

Galli, F., Lagioia, F., y Sartor, G. (2022). Consent to Targeted Advertising. *European Business Law Review*, *33*(4), 485-512. https://doi.org/10.54648/EULR2022023

García-Marzá, D., y Calvo, P. (2022). Democracia algorítmica: ¿un nuevo cambio estructural de la opinión pública? *Isegoría*, (67), e17. https://doi.org/10.3989/ISEGORIA.2022.67.17

García-Marzá, D., y Calvo, P. (2024). *Algorithmic democracy: A critical perspective from deliberative democracy*. Cham: Springer.

Garcia, D., y Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, *67*, 69-74. https://doi.org/10.1016/j.paid.2013.10.001

Gerber, A. S., Huber, G. A., Doherty, D., y Dowling, C. M. (2011). The Big Five Personality Traits in the Political Arena. *Annual Review of Political Science*, *14*, 265-287. https://doi.org/10.1146/ANNUREV-POLISCI-051010-111659

Gerodimos, R., y Justinussen, J. (2015). Obama's 2012 Facebook Campaign: Political Communication in the Age of the Like Button. *Journal of Information Technology & Politics*, *12*(2), 113-132. https://doi.org/10.1080/19331681.2014.982266

González Moraga, F. R. (2015). La tríada oscura de la personalidad: maquiavelismo, narcisismo y psicopatía. *Revista Criminalidad*, *57*(2), 253-265. Recuperado de https://dialnet.unirioja.es/servlet/articulo?codigo=5456799

Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. En M. Seelinger & S. Sevignani (Eds.), *Ein neuer Strukturwandel der Öffentlichkeit?* (pp. 470-500). Baden-Baden: Nomos.

Haig, Z., y Hajdu, V. (2017). New Ways in the Cognitive Dimension of Information Operations. *Land Forces Academy Review*, *22*(2), 94-102. https://doi.org/10.1515/raft-2017-0013

Han, B. C. (2021). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder.

Helmus, T. C. (2022). *Artificial Intelligence, Deepfakes, and Disinformation*. RAND Corporation. https://doi.org/10.7249/PEA1043-1

Hersh, E. D. (2015). *Hacking the Electorate: How Campaigns Perceive Voters* . Cambridge: Cambridge University Press.

Hoffman, S., y Attrill, N. (2021). *Mapping China's Tech Giants: Supply chains and the global data collection ecosystem* (N.º 45/2021). The Australian Strategic Policy Institute. Recuperado de https://ad-aspi.s3.ap-southeast-2.amazonaws.com/2021-06/Supply chains.pdf?VersionId=56J_tt8xYXYvsMuhriQt5dSsr92ADaZH

Howard, P. N. (2020). *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale Univeristy Press.

Issenberg, S. (2012). *The victory lab: the secret science of winning campaigns*. New York: Broadway Books.

Iyer, P., Riedl, M. J., Trauthig, I. K., y Woolley, S. (2021). *Location-based targeting: history, usage, and related concerns*. University of Texas at Austin. Center for Media Engagement. Recuperado de https://mediaengagement.org/research/location-based-targeting-history-usage-and-related-concerns/

Jamieson, K. H. (2018). *Cyberwar: How Russian hackers and trolls helped elect a president: What we don't, can't, and do know*. Oxford: Oxford University Press.

Juárez Ramos, V. (2019). *Analyzing the Role of Cognitive Biases in the Decision-Making Process*. Hershey: IGI Global.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux Inc.

Kaiser, B. (2019). *Targeted: My Inside Story of Cambridge Analytica and How Trump, Brexit and Facebook Broke Democracy*. London: HarperCollins.

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific Reports*, *11*(1), 100. https://doi.org/10.1038/s41598-020-79310-1

Kosinski, M., Stillwell, D., y Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802-5805. https://doi.org/10.1073/pnas.1218772110

Lastra-Anadón, C., y Rubio, D. (2020). *European Tech Insights 2020*. IE Center for the Governance of Change (CGC). Recuperado de https://docs.ie.edu/cgc/CGC-European-Tech-Insights-2020.pdf

Llaneza, P. (2019). *Datanomics: Todos los datos personales que das sin darte cuenta y todo lo que las empresas hacen con ellos*. Barcelona: Deusto.

Matz, S. C., Kosinski, M., Nave, G., y Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12714-12719. https://doi.org/10.1073/PNAS.1710966114/SUPPL_FILE/PNAS.1710966114.SAPP.PDF

Mayer-Schönberger, V., y Cukier, K. (2013). *Big data: La revolución de los datos masivos*. Madrid: Turner Publicaciones.

Mercier, H., y Sperber, D. (2017). *The Enigma of Reason*. Cambridge: Harvard University Press.

Millière, R. (2022). Deep learning and synthetic media. *Synthese*, *200*(3), 1-27. https://doi.org/10.1007/S11229-022-03739-2/FIGURES/6

Moriyama, T. (2022). *Empire of Direct Mail: How Conservative Marketing Persuaded Voters and Transformed the Grassroots*. Kansas: University Press of Kansas

Nave, G., Greenberg, D. M., Kosinski, M., Stillwell, D., y Rentfrow, J. (2018). Musical Preferences Predict Personality: Evidence from Active Listening and Facebook Likes. *Pyschological Science*, *29*(7), 1145-1158. https://doi.org/10.1177/0956797618761659

Nightingale, S. J., y Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8), e2120481119. https://doi.org/10.1073/PNAS.2120481119/ASSET/E74865F1-3BC4-4BEC-8325-DEB222AE2CB4/ASSETS/IMAGES/LARGE/PNAS.2120481119FIG04.JPG

Perez Colome, J., y Ayuso, S. (22 de mayo de 2023). Irlanda impone a Meta una multa de 1.200 millones de euros, la mayor sanción europea por infracción de privacidad. *El País*. Recuperado de https://elpais.com/tecnologia/2023-05-22/irlanda-impone-a-meta-una-multa-de-1200-millones-de-euros-la-mayor-sancion-europea-por-infraccion-de-privacidad.html

Rosenberg, M., Confessore, N., y Cadwaladr, C. (17 de marzo de 2018). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. Recuperado de https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Sanborn, F. W., y Harris, R. J. (2018). *A cognitive psychology of mass communication*. New York: Routledge.

Saura García, C. (2023). El big data en los procesos políticos: hacia una democracia de la vigilancia. *Revista de filosofía*, *80*, 215-232. https://doi.org/10.4067/S0718-43602023000100215

Saura García, C. (2024). Digital expansionism and big tech companies: consequences in democracies of the European Union. *Humanities and Social Sciences Communications*, *11*(448), 1-8. https://doi.org/10.1057/s41599-024-02924-7

Schick, N. (2020). *Deep Fakes and the Infocalypse : What You Urgently Need To Know*. London: Monorary.

Schumpeter, J. A. (1958). *Capitalismo, socialismo y democracia*. Madrid : Aguilar.

Singer, N. (15 de septiembre de 2022). This Ad's for You (Not Your Neighbor). *The New York Times*. Recuperado de https://www.nytimes.com/2022/09/15/business/custom-political-ads.html

Solove, D. J. (2021). The Myth of the Privacy Paradox. *George Washington Law Review*, *89*(1), 1-51. https://doi.org/10.2139/SSRN.3536265

Thompson, S. A., y Warzel, C. (19 de diciembre de 2019). Twelve Million Phones, One Dataset, Zero Privacy. *The New York Times*. Recuperado de https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html

Turow, J. (2021). *The voice catchers : how marketers listen in to exploit your feelings, your privacy, and your wallet*. New Haven: Yale University Press.

Varoufakis, Y. (2023). *Technofeudalism: What Killed Capitalism*. London: Random House.

Woolley, S. (2023). *Manufacturing consensus : understanding propaganda in the era of automation and anonymity*. New Haven: Yale University Press.

Woolley, S., y Gursky, J. (21 de junio de 2020). The Trump 2020 app is a voter surveillance tool of extraordinary power. *MIT Technology Review*. Recuperado de https://www.technologyreview.com/2020/06/21/1004228/trumps-data-hungry-invasive-app-is-a-voter-surveillance-tool-of-extraordinary-scope/

Woolley, S., & Howard, P. N. (Eds.). (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford: Oxford University Press.

Wylie, C. (2019). *Mindf*ck. Inside Cambridge Analytica's Plot to Break the World.* London: Profile Books.

Youyou, W., Kosinski, M., y Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040. https://doi.org/10.1073/PNAS.1418680112

Zakrzewski, C. (4 de julio de 2023). Judge blocks U.S. officials from tech contacts in First Amendment case. *The Washington Post*. Recuperado de https://www.msn.com/en-us/news/politics/judge-blocks-us-officials-from-tech-contacts-in-first-amendment-case/ar-AA1dq6Cj

Zewe, A. (17 de enero de 2020). Imperiled information: Students find website data leaks pose greater risks than most people realize. *Harvard John A. Paulson School of Engineering and Applied Sciences*. Recuperado de https://seas.harvard.edu/news/2020/01/imperiled-information

Zuboff, S. (2020). *La era del capitalismo de la vigilancia: la lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós.

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 91-117
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.612061

Uncommon ground y pluralidad de actos de habla en polílogos online

Uncommon ground and plurality of speech acts in online polylogues

*CATARINA MACHIONI SPAGNOL**

Resumen: ¿Cómo los colectivos argumentan y por qué hay tantos desacuerdos? Tal como explican Lewiński y Aakhus, el modelo tradicional de argumentación, que reduce las múltiples posiciones expresadas en las redes sociales a una dicotomía de proponente versus oponente, es insuficiente para solucionar racionalmente nuestros desacuerdos porque no captura la realidad polilógica presente en las comunicaciones online a gran escala. Utilizando ejemplos de discusión en *X* (anteriormente conocida como Twitter) debido a su capacidad para generar interacciones públicas amplias y diversas, en este artículo propongo que las múltiples posiciones reflejan la diversidad de creencia e intenciones de los participantes, organizándose en *clusters* de actos de habla que emergen del *uncommon ground* (falta de información compartida sobre el mundo).
Palabras clave: polílogo, comunicación *online*, argumentación, actos de habla, desacuerdo

Abstract: How do collectives argue and why are there so many disagreements? As Lewiński and Aakhus explain, the traditional model of argumentation, which reduces the multiple positions expressed on social media to a dichotomy of proponent versus opponent, is insufficient for rationally resolving our disagreements because it does not capture the polylogical reality present in large-scale *online* communications. Using examples of discussions on *X* (formerly known as Twitter) due to its capacity to generate broad and diverse public interactions, this article proposes that the multiple positions reflect the diversity of beliefs and intentions among participants, organizing into clusters of speech acts that emerge from the uncommon ground (lack of shared information about the world)
Keywords: polylogue, online communication, argumentation, speech acts, disagreements

---

Recibido: 12/04/2024. Aceptado: 27/06/2024.

* UNED, Máster Universitario de Filosofía Teórica y Práctica (Especialidad Lógica, Historia y Filosofía de la Ciencia). Este trabajo ha sido desarrollado bajo la orientación de Maria Cristina Corredor Lanas como actividad práctica para la asignatura «Temas de Pragmática, Argumentación y Actos de Habla». Sus líneas de investigación giran principalmente en torno a: 1) Filosofía del lenguaje, teorías del significado, lenguaje y sociedad, 2) Perspectivas de cambios conceptuales, sociales y políticos desde la filosofía del lenguaje y 3) Prácticas colectivas de argumentación y comunicación relacionadas con los desacuerdos. Correo electrónico: catarinamachionispagnol@gmail.com

1. Introducción

Afirma María José Frápolli Sanz (2024) que somos animales comunicacionales. En la misma línea, Dennett sostiene que «no importa lo distintos que seamos los unos de los otros, diseminados como estamos por todo el globo, pues podemos explorar nuestras diferencias y comunicarnos acerca de ellas» (Dennett, 2004, p. 18). Y para negociar nuestras diferencias comunicacionales, hemos desarrollado nuevas tecnologías de comunicación. Históricamente, la difusión masiva de estas nuevas tecnologías a finales del siglo XX y principios del XXI proporcionó nuevas formas de interacción, facilitando así este aspecto comunicativo humano (Lewiński y Aakhus, 2023).

Actualmente, la comunicación *online* es un fenómeno complejo, caracterizado por una dinámica abierta, difícil de controlar y que involucra a un número amplio de personas. Esta difusión masiva de la comunicación presenta desafíos. Por un lado, la tecnología informacional permite una mayor interacción entre las personas, ampliando los procesos deliberativos y la amplia participación en el debate público.[1] Por otro lado, ha producido nuevas formas de desacuerdo, generando el problema de la falta de consenso a gran escala (Lewiński y Aakhus, 2023).

En este contexto no ideal de la comunicación *online*, el principal problema radica, según autores como Innocenti (2022), Aakhus y Lewiński (2017) y Lewiński et al. (2023), en la dificultad de encajar la pluralidad conversacional en la perspectiva tradicional de protagonista versus el antagonista. El modelo diádico de la teoría de la argumentación y de los actos de habla no logra establecer estrategias para la resolución racional de nuestros desacuerdos, especialmente aquellos que involucran a múltiples participantes y generan una pluralidad de actos de habla (p. 188-189).[2] Por lo tanto, la propagación de los nuevos medios de interacción ha resaltado los límites del análisis basado en las formas dialógicas tradicionales (Lewiński y Aakhus, 2023, p. 7).

Para contrarrestar el método dialéctico de análisis, surge la noción de polílogo. Para Kerbrat-Orecchioni (2004), un polílogo puede ser definido como una interacción entre múltiples participantes, de manera que toda situación comunicativa que reúne desde cuatro a un número infinito de integrantes puede ser considerada polilógica (p. 3-4). Para Aakhus y Lewiński (2017), un polílogo puede ser definido como una interacción conversacional argumentativa que involucra (i) los múltiples participantes (incluyendo grupos diversos), reivindicando (ii) las múltiples posiciones argumentativas a través de (iii) múltiples lugares (p. 181).

Acepto esta segunda definición y abro aquí un breve paréntesis para dar unas pinceladas a cada una de estas nociones[3]. Sin buscar una simplificación excesiva, se puede decir que los múltiples participantes son los agentes comunicativos involucrados en una determinada inte-

1 Por ejemplo, Landemore (2021) propone que las tecnologías digitales pueden actuar como mediadoras en la transición hacia una democracia abierta y deliberativa. Según ella, esto sería posible mediante un rediseño de las plataformas para mejorar los problemas relacionados con la manipulación de información y garantizar el derecho a la privacidad de los usuarios.

2 Esta articulación fue motivada por comentarios de Bruno Ramos Mendonça, los cuales agradezco.

3 Comparando las dos definiciones, considero que la segunda es más amplia, ya que también involucra las posiciones y los lugares, lo que permite acomodar la realidad polilógica de manera más adecuada.

racción (Lewiński y Aakhus, 2023, p. 78) que expresan, por ejemplo, duda, preocupación, oposición, contribuyendo así a la pluralidad de la interacción, y que influyen en el diseño estratégico de la argumentación (Palmieri y Mazzali-Lurati, 2016, p.472). Las múltiples posiciones son las proposiciones compartidas por los participantes y que surgen de diversas maneras en un polílogo. No corresponden a los lados de una discusión (Lewiński y Aakhus, 2023, p. 77), sino a los puntos de vista que suelen estar en relación con otras posiciones en relación de oposición, complementariedad, etc. Como se observará, funcionan como *clusters* para los actos de habla. Por último, los múltiples lugares equivalen al momento y al lugar donde los agentes comunicativos se expresan (p. 81). El lugar actúa como el contenedor de la interacción, por ejemplo, un tribunal, una clase, un sitio web o una red social. En resumen, los múltiples lugares se refieren al contexto de las conversaciones. Una manera más precisa de entender la dinámica del polílogo es tener en mente la relación *who-what-where* (Aakhus y Lewiński, 2017, p.199), como ilustra la imagen a continuación. Y con estas pinceladas, cierro el paréntesis.

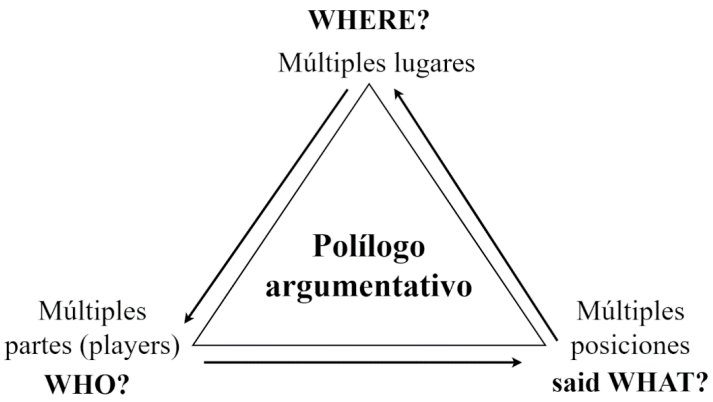

Figura 1. El modelo tripartito de un polílogo

A la luz de los problemas enunciados y considerando la idea de los polílogos como realidad comunicacional, autores como Musi y Aakhus (2018), Lewiński y Aakhus (2014) e Innocenti (2022) defienden la apertura a nuevas formas de observaciones y análisis de esta realidad comunicacional polilógica. Por ejemplo, Lewiński y Aakhus (2014) e Innocenti señalan la necesidad de mirar hacia los polílogos desde diferentes puntos de partida, considerando principalmente la característica de la falta de los compromisos compartidos y el no cumplimento de aspectos normativos en las conversaciones a gran escala, lo que puede afectar el desarrollo de la argumentación. Musi y Aakhus (2018) se enfocan en la necesidad de observar los patrones conversacionales y argumentativos, lo que implica dos desafíos empíricos: (i) detectar las características de la argumentación en las redes sociales y (ii) distinguir los argumentos usados para la manutención del desacuerdo de los utilizados para otros fines (p. 399).

Este artículo es el resultado de un estudio preliminar. Mi objetivo es contribuir a la identificación de cómo los colectivos argumentan; una identificación que ayude a observar patrones en la comunicación *online*. Partiendo de la aceptación de que los modelos tradicionales de argumentación no dan cuenta de acomodar la realidad polilógica de la comunicación *online* a gran escala, a lo largo del artículo plantearé el conjunto de tesis que justificaré a lo largo del artículo:

1. La comunicación *online* a gran escala (CLGE) es una realidad polilógica (RP).[4]
2. Una RP emerge de múltiples participantes (MPA), múltiples posiciones (MPO) y múltiples lugares (ML).
3. Existen casos en los que una RP que emerge de MPO se organiza como *clusters* de actos de habla (AH).
4. Los AH expresan la diversidad de creencias e intenciones del conjunto de hablantes que participan de la RP.
5. Las creencias e intenciones del conjunto de hablantes que participan de la RP emergen de un *uncommon ground* (UG).
6. Por tanto, la identificación de las MPO como *clusters* de AH contribuye a la organización de la RP.

Para analizar estas conversaciones a gran escala, me enfocaré en identificar los patrones de los actos de habla interactivos existentes en las MPO, específicamente en *X* (anteriormente Twitter). Presentaré tres ejemplos de interacción polilógica en *X*, donde los participantes expresan sus posiciones sobre el reciente conflicto Israel-Palestina después de la ofensiva terrorista de Hamás el 7 de octubre de 2023. La elección de *X* como plataforma se justifica desde la perspectiva de Sara Greco (2023), quien describe esta red como un espacio de argumentación polilógica que abre sub-discusiones cuestionando hechos, valores y proposiciones de conocimiento en respuesta a argumentos de otros lugares (p. 2). Aunque las interacciones en *X* son entimemáticas (p.17), se consideran partes de una discusión argumentativa más amplia y transversal (p.4). Estoy de acuerdo con Sara Greco y considero que esta característica refuerza la conveniencia de *X* para el análisis de interacciones polilógicas *online*.

La perspectiva de Greco se complementa con los hallazgos de Nancy K. Baym (2015), quien destaca algunas peculiaridades distintivas de *X*. Baym (2015) destaca que *X* se diferencia de otras plataformas por fomentar la interactividad y alcanzar grandes audiencias mediante la replicabilidad y el intercambio rápido de información, facilitando así la organización de movimientos. Sin embargo, también es conocida por el fenómeno de la polarización, que dificulta el consenso entre los hablantes (p.92).[5] Además, Baym señala

4 Esta nota introduce términos fundamentales que serán explicados de manera adecuada a lo largo del artículo. La realidad polilógica (RP) es una situación comunicacional que involucra múltiples participantes, posiciones y lugares. Las posiciones (MPO) son las diversas proposiciones expresadas durante el debate, los (MPA) son los agentes comunicativos que intervienen en la interacción y los múltiples lugares (ML) son los espacios digitales que facilitan la interacción, por ejemplo, una red social específica. El *uncommon ground* (UG) corresponde a la falta de información y creencias compartidas entre los hablantes, influenciando la dinámica de las interacciones.

5 Entiendo que, para Baym, la polarización puede ser comprendida como el fenómeno que captura las diferencias a través de la división de opiniones diversas como divisibles en grupos extremos y homogéneos en términos

que la estructura de *X* promueve relaciones sociales superficiales y facilita la difusión de información falsa, ya que permite simultáneamente el anonimato y la visibilidad pública de los usuarios[6].

Para concluir, se demostrará que la organización de los actos de habla en *clusters* proposicionales es una herramienta valiosa para gestionar los desacuerdos en las redes sociales. Al reconocer la comunicación *online* a gran escala como no-ideal y adoptar la idea de UG como punto de partida, se puede adaptar el modelo de polílogo para analizar y mejorar nuestras interacciones argumentativas en estos entornos complejos.

2. Comunicación *online* y uncommon ground

La comunicación *online* a larga escala (CLGE) se caracteriza, en primer lugar y de manera intuitiva, por la mediación de un ordenador. Según Michel Marcoccia (2004), esta mediación implica conversaciones escritas (p. 116) que integran el lenguaje escrito en la cotidianidad (Borg y Connolly, 2022, p.8). Además, Marcoccia resalta el aspecto multimodal de estas interacciones, donde se incorporan no solo las palabras escritas, sino también imágenes y vídeos. Otra característica importante es la asincronía, que permite a los participantes comunicarse en momentos diferentes, generando así lagunas, discontinuidad y superposiciones debido a publicaciones simultáneas. Estos elementos añaden complejidad a la estructura conversacional. Además, Marcoccia (2004) destaca el carácter (y acceso) público a estas conversaciones. Estas características, sostiene este autor, evidencian fenómenos como la cacofonía y la fragmentación de las conversaciones. Otro factor relevante en la comunicación *online* es el efecto a gran escala que permite una amplia y sistémica difusión de la conversación.

Además, las redes sociales facilitan diversas modalidades comunicativas que expresan deseos, emociones y otros aspectos, por ejemplo, psicológicos y estéticos. Mclver Lopes (2014) sostiene que, como ciudadanos, formamos opiniones sobre debates sociales basados también en nuestras percepciones estéticas. En la dimensión estética de la comunicación, la simetría, como propiedad del sistema comunicativo de valores en las interacciones mediante imágenes digitales, juega un papel significativo en el convencimiento y refuerzo de creencias (Spagnol, 2024)[7]. La tecnología también desempeña un rol crucial en esta dinámica comunicativa. Por ejemplo, Danaher (2024) argumenta sobre el papel de los influenciadores digitales en nuestras percepciones y valores, lo que incide en cambios en la moral social y las responsabilidades dentro de las relaciones. Wolff (2019) argumenta que el funcionamiento algorítmico configura las interrelaciones entre los usuarios, generando encuentros que facilitan nuevas formas de conocimiento y acción. Esto fortalece la visibilidad y fomenta

---

de pensamientos que se oponen entre sí. Parece, pues, que Baym articula esta polarización dentro del modelo tradicional que reduce las divergencias entre un bando proponente versus un bando oponente. En contraste, lo que propongo se distingue de la propuesta de Baym y se alinea con el modelo polilógico. Este enfoque permite abrazar la diferencia en términos distintos a «nosotros contra ellos», por ejemplo.

6 La inclusión de este párrafo en esta sección ha sido motivada por comentarios de un revisor anónimo, a quien agradezco.

7 Todavía sin publicar.

la participación en publicaciones específicas que incentivan el interés colectivo en distintos temas sociales[8].

Por otra parte, según explica Robert Stalnaker (2002), el *common ground* en una conversación consiste en la formulación de proposiciones que los participantes asumen mutuamente, dándolas por sentadas (Macagno y Capone, 2016, p. 152). Sin incurrir en simplificaciones excesivas, esta noción puede definirse como el conjunto de conocimientos y creencias compartidas que los participantes presuponen durante una interacción comunicativa. Stalnaker (2002) defiende que el *common ground* relacionado con una proposición *P* se establece si todos los participantes aceptan que *P* y todos creen que todos aceptan *P* (p. 716). En resumen, el *common ground* no es solo un conjunto de creencias compartidas entre los hablantes, sino la creencia de que comparten estas creencias (Lewiński y Aakhus, 2023, p. 103). Y esta creencia sirve como base para la comunicación efectiva.

En contraste, en situaciones de «contextos defectuosos», los participantes no comparten las mismas presuposiciones, lo que requiere un proceso de acomodación para resolver las diferencias (Stalnaker, 2002, p. 717). Este fenómeno de acomodación implica ajustes en la información para facilitar la comprensión mutua y la cooperación comunicativa (p. 711). Stalnaker también reconoce que no todos los contextos defectuosos pueden corregirse fácilmente, y en algunos casos, el *common ground* simplemente no se establece. Y esto resulta en una condición comunicativa no ideal.

Volvamos a las características de la comunicación *online* propuestas por Marcoccia (2004), principalmente, el aspecto multimodal y la asincronía. En este tipo de comunicación, la participación se abre a múltiples participantes, generando cacofonía y fragmentación de la conversación. Este fenómeno de fragmentación proporciona una condición no ideal de comunicación, configurando un contexto comunicativo defectuoso. La fragmentación dificulta que los hablantes sepan, de manera evolutiva y conjunta, qué se ha dicho, quién lo ha dicho y por qué lo ha dicho. Este registro conversacional es crucial para la construcción conjunta de la conversación.[9] En la dinámica de las redes, por ejemplo, *X*, la construcción conjunta está fragmentada; aspecto que dificulta el estándar de precisión. Cuando no se puede construir conjuntamente el registro conversacional, la noción de *common ground* planteada por Stalnaker no se da, ya que no se construye conjuntamente la información de base para la conversación: los hablantes aterrizan sin aviso en una interacción asíncrona y fragmentada. Con estas consideraciones en mente, la noción de *uncommon ground* (UG) que propongo aquí designa la falta de la creencia en las creencias compartidas entre los participantes.

Para ilustrarlo, cabe iimaginar un contexto deliberativo donde los participantes deben debatir sobre dónde los «hijos de Fred» pueden dormir. En una conversación no fragmentada y con un *common ground*, si un hablante menciona que «todos los hijos de Fred están dormidos», se supone que todos los hablantes comparten la creencia de que todos tienen la creencia de que «Fred tiene hijos»[10]. Continuando la analogía, en una estructura de conversa-

8 La inclusión de este párrafo en esta sección ha sido motivada por comentarios de revisores anónimos, a quienes agradezco.

9 El concepto de «registro conversacional» fue planteado por David Lewis (1979). Según él, una conversación dispone de un registro que incluye información sobre lo que ha sido dicho, las creencias de los participantes, las intenciones comunicativas y las presuposiciones compartidas.

10 El ejemplo utilizado fue inspirado en el artículo *Scorekeeping in a Language Game* de David Lewis (1979).

ción fragmentada y asíncrona tal como se da en *X*, los hablantes no sabrían quién es «Fred», si tiene hijos, o cuántos hijos tiene, etc., pero quieren participar en la decisión sobre dónde pueden dormir. Para ilustrar, el UG sería la falta de la creencia en la creencia compartida de que «Fred tiene hijos»[11].

Planteo que el UG es la falta de información compartida sobre el mundo. Es más, sugiero que el UG incorpora la diversidad de creencias e intenciones individuales de los hablantes en la conversación, lo que genera una complejidad psicológica y epistémica que lleva a contradicciones y confusiones (Williamson, 2002, cap. 4.5). En el contexto de la comunicación *online*, especialmente en plataformas como *X*, propongo que el UG es una condición fundamental. Y esto se debe a la naturaleza asíncrona y fragmentada de las conversaciones en redes sociales. No obstante, si queremos mantener debates democráticos en las redes, debemos considerar esta característica como parte integral, incorporándola a las conducciones de los debates en situaciones donde las argumentaciones son necesarias. Adaptando la idea de UG para el modelo de Lewiński y Aakhus (2023), se puede decir que este UG se equipara al «espacio de desacuerdo» (p. 184) existente en la realidad polilógica, configurándose como una propiedad del sistema mismo de comunicación a gran escala.

Esta idea se apoya en la perspectiva de Meijers (2007), quien argumenta que, fuera de contextos de acciones colectivas, los hablantes individuales pueden creer erróneamente que sus intenciones son compartidas (Meijers, 2007: p.100). Meijers sostiene que una condición ideal para el reconocimiento preciso de las intenciones y creencias de los hablantes requiere una concisión grupal y un acceso epistémico que garantice la transmisión precisa de la intención ilocutiva. Además, la razón para realizar el acto de habla debería captar completamente el interés común y no solo del hablante. La idea de UG también es compatible con la idea de la comunicación *online* como un sistema social emergente. Para Prigogine y Stengers (2018), por ejemplo, los sistemas sociales están en constante estado de cambio y evolución, y son vulnerables al desequilibrio. En este caso, el modelo basado en equilibrio equivaldría al modelo tradicional de argumentación, que visualiza condiciones y criterios ideales de desacuerdos, lo cual no ocurre en la comunicación *online* a gran escala. Lo que tenemos es una condición no ideal de comunicabilidad. Afirmar que un sistema como el comunicacional tiene una dinámica de desequilibrio no significa que esta misma dinámica no pueda ser identificada o incluso organizada, más bien que necesita de un modelo adecuado para acomodarla.

Con estas aclaraciones en mente, volvamos a las múltiples posiciones (MPo). Las MPo son una manifestación de la diversidad de creencias e intenciones de los hablantes, las cuales se expresan en los actos de habla como fuerzas ilocutivas. Propongo que estos actos suelen organizarse de tal manera que las MPo funcionan como *clusters* que emergen del UG, lo cual se corresponde con la realidad polilógica. Aakhus y Lewiński (2017) defienden que, en contextos de desacuerdo, inevitablemente habrá múltiples partes involucradas y una diversidad de posiciones y por ende, un polílogo. Autores como Innocenti (2022) y Aakhus y Lewiński (2017) argumentan que los polílogos *online* se caracterizan por la controversia y la falta de normatividad presente en teorías clásicas de la argumentación y los actos de habla. Lewiński (2021) argumenta que las interacciones polilógicas pueden

11 Agradezco a Bruno Ramos Mendonça por haber llamado mi atención sobre el hecho de que la noción de UG estaba oscura y necesitaba clarificaciones adicionales.

tener múltiples «*shared grounds*» entre los participantes (p. 436), lo cual refuerza la idea de UG como la base de estas interacciones. Si se aceptan estas justificaciones, también es plausible afirmar que el UG establece un punto de partida crucial para entender y analizar los actos de habla en los polílogos *online*, lo que contribuye a comprender la complejidad de las interacciones comunicativas en los entornos digitales y a abordar la gestión de los desacuerdos en estos contextos.

Conviene mencionar que considerar el UG como punto de partida para el análisis de los actos de habla en los entornos digitales no implica afirmar que todas las conversaciones en las redes sociales lleven a desacuerdos o estén orientadas al debate argumentativo. Es posible tener conversaciones agradables o expresar acuerdos en estas redes. Sin embargo, algunas de las conversaciones que se desarrollan, por ejemplo, en *X*, cuando despiertan el interés público fomentan el debate e influyen en la toma las decisiones. En estas situaciones de interés general en temas públicos, el debate acoge una diversidad de personas y luego, conlleva a una pluralidad de actos de habla. Si, según Stalnaker (2002: p. 708), un acto de habla introduce una creencia en la conversación, la organización en *clusters* posicionales transformados en objeto de organización y estructurados en el modelo polilógico facilitan la comprensión del debate. Al analizar las conversaciones, no solo se debe presuponer un UG, sino reconocerlo como base subyacente.

3. Múltiples posiciones y pluralidad de actos de habla

Empecemos esta sección justificando por qué el modelo tradicional es inadecuado para acomodar las múltiples posiciones (MPo). Lewiński y Aakhus (2014) sugieren dos enfoques posibles para analizar estas posiciones dentro del modelo tradicional: (i) reducir las MPo a dos bandos, uno a favor y otro en contra y (ii) adoptar la perspectiva retórica que considera las MPo como si fuera una audiencia universal (p. 166). No obstante, ambos enfoques enfrentan limitaciones significativas. El primero simplifica demasiado la diversidad de matices presentes en un polílogo al enmarcar las posiciones en roles tradicionales de antagonista y protagonista (p. 171). Mientras tanto, la perspectiva retórica, aunque reconozca la heterogeneidad de las posiciones, no captura la dinámica interactiva multidireccional del polílogo (p. 170). Parece, pues, que reducir la argumentación a un modelo bilateral es un obstáculo para comprender la complejidad de las conversaciones en entornos digitales. Para resolver estas limitaciones, surge la necesidad de un enfoque que pueda capturar la complejidad inherente a las interacciones digitales. La noción de polílogo, tal como planteada por Lewiński y Aakhus (2023), surge como una estructura capaz de organizar adecuadamente las posiciones. Para una comprensión más adecuada de cómo se organizan las múltiples posiciones en la comunicación *online*, conviene considerar no solo la estructura polilógica, sino el papel que los actos de habla desempeñan en estas interacciones.

Lewiński (2021) plantea que las conversaciones *online* también no deben reducirse a un intercambio de información, sino considerarse un tipo de actividad humana donde los actos de habla desempeñan un papel fundamental en la realización de nuestros objetivos individuales y colectivos (p.422). Reconocer esto implica aceptar una presunción de racionalidad detrás de los actos de habla, los cuales contribuyen a la construcción de argumentos

y posiciones divergentes (p. 424). Además, según explican Kauffeld y Goodwin (2022, p. 2-5), la perspectiva pragmática de los AH enfatiza que su fuerza ilocutiva subyace en una presunción racional mediante el habla. Según la pragmática, la presunción contenida en el AH contribuye a la conversación, añadiendo información. En este caso, los participantes se encuentran con la intención que el hablante busca lograr al realizar un AH específico. Cuando se combina con el contenido semántico del acto, que corresponde la base literal y conceptual del mensaje, forma un acto de habla completo y efectivo. Así, se puede decir que un AH contiene el esfuerzo comunicativo primitivo del hablante. Por ejemplo, el AH de *prometer* carga simultáneamente la conversación con una base literal y conceptual del verbo y con la fuerza ilocutiva de hacer una promesa. Si el hablante afirma algo, hace un esfuerzo comunicativo para transmitir una proposición como verdadera; por ejemplo, si emite un acto de habla de acusación, está imponiendo la obligación de explicar la conducta acusada, y así sucesivamente[12].

Otro aspecto relevante en la teoría de los actos de habla, como explica Grice (1975), es la importancia de la intención del hablante en inducir creencias en el oyente y en ser reconocido como tal (p. 383). Este proceso implica un entendimiento normativo entre hablante y oyente, donde la efectividad de la comunicación depende de la aceptación de las intenciones comunicativas. Cristina Corredor (2020),[13] amplía este punto al distinguir entre dos modos de razonamiento (asociados con las nociones fregeanas y griceanas de normatividad inferencial): uno automático e intuitivo, y otro controlado y reflexivo, ambos fundamentales para entender la dinámica inferencial y normativa en las interacciones argumentativas (p. 46-47). Desde esta perspectiva, la argumentación no solo cumple una función comunicativa, sino también una función epistémica crucial para la comprensión y evaluación de los argumentos en disputa. Comprender que una realidad polilógica, emergente de MPo, se organiza como *clusters* de actos de habla es crucial. Estos *clusters* expresan la diversidad de creencias e intenciones de los participantes y emergen de un UG. Reducir esta pluralidad a una simple dicotomía de proponente versus oponente es, según Lewiński y Aakhus (2023), un error falaz. Si la estructura conversacional es polilógica, el modelo de análisis también debe serlo.

Para ahondar en este aspecto, consideremos el siguiente argumento. Según Lewiński y Aakhus (2023), el problema de pensar la realidad polilógica en términos de proponente versus oponente conduce a la falacia del falso dilema (p. 184). Si el UG incorpora la diversidad de creencias e intenciones de los hablantes que encuentran en los actos de habla su vía para la expresión como fuerzas ilocutivas, el modelo polilógico necesita acomodar estas diferencias, manteniendo el espacio de desacuerdo. Como será demostrado, el polílogo, a diferencia del modelo diádico de proponente-oponente, amplía la perspectiva para incluir

12 Para entender mejor la pluralidad de actos de habla, se presenta la siguiente taxonomía: (1) Asertivo (*assertive*): Afirmación o negación de proposiciones para transmitir información que se presume verdadera; (2) Acusatorio (*accusing*): Imputación de culpa o responsabilidad a alguien; (3) Directivo (*directive*): Busca que el oyente realice una acción; (4) Interrogativo (*interrogative*): Busca obtener información a través de una pregunta; (5) Expresivo (*expresive*): Comunicación de sentimientos o emociones sobre un hecho; (6) Exhortación (*exhortation*): Incita a hacer o dejar de hacer algo; (7) *Asesoramiento* (*Advising*): Recomendación o consejo; (8) Declarativo (*declarative*): Propósito de provocar un cambio en la situación; (9) Comisivo (*commissive*): Comprometimiento a una acción futura.

13 La referencia para la distinción que desarrolla Cristina Corredor es la obra de Tversky y Kahneman (1974).

múltiples posiciones. Además, Lewiński y Aakhus argumentan que las diferencias de opiniones y razones no son fallas en la comunicación, sino parte del proceso deliberativo. La noción de polílogo no descarta las condiciones de verdad, sino que plantea que estas no necesitan estar vinculadas únicamente a dos representantes[14]. Las partes involucradas (aunque indirectamente) en una situación interactiva deliberativa actúan según sus propias razones (aunque esta acción se reduzca a la emisión de una opinión). En este sentido, la aceptación de las razones por parte de unos depende de la capacidad de justificación por parte de otros. Importa mencionar que la noción de polílogo no descarta las condiciones de verdad presentes en la conversación, sino que plantea que esta condición no necesita estar vinculada únicamente a dos representantes (proponente versus oponente).

Una vez entendida la necesidad de un enfoque más inclusivo, surge la pregunta sobre cómo la noción de UG puede proporcionar una estructura útil para analizar estos contextos. Adoptar la noción de UG para analizar los actos de habla en los polílogos *online* es plausible porque revela cómo estas interacciones argumentativas no cumplen con las condiciones normativas ni con la función epistémica necesaria para la comunicación efectiva. Considero que este enfoque contribuye a capturar la complejidad de las interacciones actuales y a pensar nuevas formas para desarrollar la gestión de nuestros desacuerdos en los entornos digitales.

4. Las múltiples posiciones como *clusters* de la pluralidad de actos de habla

Lewiński (2021) defiende la existencia de un pluralismo ilocutivo en los polílogos argumentativos. Abro un paréntesis para introducir algunas clarificaciones importantes. Según van Eemeren y Grootendorst (2003), nuestras expresiones verbales se transforman en objetos de interés para la teoría de la argumentación cuando las utilizamos en las situaciones en las que queremos lograr determinados objetivos. Por supuesto, no todos nuestros actos de habla en las redes sociales son relevantes como objetos de investigación en este sentido. Los actos de habla que interesan para los propósitos de un análisis de este tipo son aquellos que, como explican Eemeren y Grootendorst, «expresan una postura que complementa, niega o afirma una proposición, dejando claro lo que el hablante defiende, o aquellos cuya premisa del razonamiento subyacente a la argumentación que queda implícita» (p. 3). Con esto, cierro el paréntesis.

Para comprender la relevancia de estos actos de habla en los contextos digitales, es necesario explorar cómo se manifiestan en situaciones concretas. En esta parte del artículo, mi objetivo es identificar la pluralidad de actos de habla que constituyen las MPo. Esto contribuirá a la comprensión de que, en algunos casos, la realidad polilógica que emerge

14 Esta nota representa un intento de ampliar las perspectivas interdisciplinares que la adopción del modelo polilógico podría ofrecer. Por ejemplo, en la lógica clásica, el cuadrado de oposiciones organiza las proposiciones mostrando sus relaciones de verdad: contrariedad, subcontrariedad, contradicción y subalternación. Bruno Ramos Mendonça (en conversación privada) sugirió que el multinivel comunicacional de la realidad polilógica y su multiplicidad de contextos también podría ser abordada desde la perspectiva de los diagramas de Venn. Según Ramos Mendonça (2012), estos diagramas facilitarían la visualización de proposiciones y sus relaciones en contextos variados. El aspecto dinámico y adaptable de estos diagramas permitiría representar y comunicar complejidades lógicas en debates y análisis interdisciplinarios, lo que permitiría identificar cómo las proposiciones y sus condiciones de verdad se interrelacionan en diferentes niveles y contextos.

de MPO es organizable como *clusters* de AH que expresan la diversidad de creencias e intenciones de los participantes, resultantes del UG. Para lograr esto, me centraré en los polílogos en la red social *X* y utilizaré ejemplos recientes de interacciones relacionadas con el conflicto Israel-Palestina, específicamente después de la ofensiva de Hamás contra Israel el 7 de octubre de 2023.

Antes de pasar a estos ejemplos, conviene aclarar ciertos aspectos metodológicos y teóricos.[15] En primer lugar, la comunicación en *X* se realiza principalmente de manera escrita. Según Clark (1996) la palabra escrita tiene el mismo estatus de relevancia que la comunicación oral respecto a la transmisión de significado en el contexto de uso. Clark afirma que ambas las modalidades comunicativas cumplen los principios pragmáticos necesarios para la interacción humana. En segundo lugar, aunque son importantes, esta primera etapa excluye aspectos como la manipulación de la opinión pública o el uso de algoritmos por parte de *X*.[16] En tercer lugar, autores como Neri Marsili (2020) y Emanuele Arielli (2018) consideran el «*repost*» como un acto de habla ostensivo. Aunque estoy de acuerdo con ellos, no se prestó atención a este tipo de acto de habla por la misma razón de que no cambia la dinámica central. Por último, la metodología utilizada consistió en buscar el término «*Israel-Palestina*» en la plataforma *X*,[17] identificando a los principales participantes y sus publicaciones relevantes sobre el conflicto. Se consideraron únicamente las publicaciones realizadas entre los días 7 y 8 de octubre de 2023 y se seleccionaron aquellas con más de tres mil respuestas y más de tres mil publicaciones compartidas. También se consideraron solo las respuestas con más de 50 «*likes*», excluyendo imágenes o vídeos y centrándose únicamente en las respuestas escritas.[18]

Estas fueron las tres preguntas que guiaron el análisis: (1) ¿Cómo pueden mapearse los actos de habla que emergen en los polílogos de *X*? (2) ¿Qué patrones, si los hay, caracterizan estas interacciones polilógicas? (3) ¿Pueden las condiciones normativas ser observadas en los polílogos *online*?

A la luz de estas aclaraciones, considero que identificar las intenciones subyacentes en la realidad polilógica de las interacciones *online* importa porque permite comprender cómo las intenciones de los hablantes en las conversaciones guían las interpretaciones y respuestas entre ellos. Un enfoque adicional que puede enriquecer este análisis es la consideración de los sesgos cognitivos y las falacias en los AH. Dado la diversidad de intenciones y creencias contenidas en los AH, organizarlos en *clusters* proposicionales permite analizar qué tipos de sesgos cognitivos y falacias interfieren en la argumentación. Battersby y Bailin (2011), por ejemplo, consideran que los sesgos influyen en nuestras percepciones, recuerdos e inter-

15 La inclusión de estas observaciones en esta sección ha sido sugerida por un revisor anónimo, a quien se le agradece su perspicaz observación.

16 Tal como lo veo, considero que, aunque estos factores sean relevantes, actos de habla escritos por *robots,* por ejemplo, no alterarían la dinámica que estoy intentando captar aquí, es decir, la fuerza ilocutiva de los actos de habla ni su inclusión en el *cluster* posicional. Por el contrario, permitirían observar cómo las automatizaciones interfieren en el debate público.

17 El resultado de la búsqueda puede ser verificado a través del siguiente enlace: «Israel-Palestina»

18 Reconozco que la metodología empleada puede tener limitaciones para un estudio más amplio destinado a identificar diversos patrones *online*. Sin embargo, considero que el enfoque aplicado ha permitido establecer indicios iniciales valiosos para investigaciones futuras. Aunque pueda requerir validación adicional, esta metodología ha permitido observar la plausibilidad del fenómeno discutido y la dinámica emergente.

pretaciones, afectando así nuestras tomas de decisiones. Por eso, estos autores defienden que comprender el ámbito práctico de las creencias actuales en un área es importante para la evaluación, especialmente en la medida en que esto determina la carga de la prueba. En la misma línea, autores como Walton (2008) e Eemeren y Houtlooser (2007) argumentan sobre la relación entre las falacias, las creencias y el comportamiento y cómo contribuyen para las decisiones. Para ellos, las falacias indican errores en el razonamiento que conlleva a conclusiones incorrectas. Las creencias facilitarían esta tendencia al error, principalmente cuando están arraigadas. En otras palabras, ellos defienden que una persona con creencias fuertes está más susceptible a razonar equivocadamente.

4.1. Ejemplo 1

El ejemplo 1 (tabla 1) describe algunas de las primeras manifestaciones públicas de personas socialmente relevantes sobre el conflicto Israel-Palestina después de la ofensiva de Hamás, como Biden, Macron, Modi, von der Leyen, etc. Dentro de la metodología aplicada, estas publicaciones tuvieran relevancia significativa y son actos de habla (AH) que pueden ser organizados en *clusters* posicionales.[19]

Posiciones	Actos de habla
La intolerancia hacia aquellos que quieren sacar ventajas sobre la situación en Israel	**(1) Joe Biden** (@POTUS), ACTO DE HABLA: «ADVISING» {Let me say this as clearly as I can. This is not a moment for any party hostile to Israel to exploit these attacks to seek advantage. My Administration's support for Israel's security is rock solid and unwavering.}
La situación de vulnerabilidad de los palestinos hacia el Estado de Israel	**(2) Ilhan Omar** (@IlhanMN), ACTO DE HABLA: ASERTIVO {Gaza's 2+ million population are mostly children, who live under blockade in what Israel's own former intelligence chief has called an open air prison. The overwhelming majority live in poverty. Many suffer lifelong psychological and physical trauma.}

19 Esta imagen no muestra todos los AH desencadenados por cada posición. Omití la amplia cadena de interacciones (las respuestas a las respuestas y así sucesivamente) únicamente por motivos de espacio y practicidad.

El rechazo al acto terrorista de Hamás y la solicitud al Estado de Israel	**(3) Emmanuel Macron** (@EmmanuelMacron), ACTO DE HABLA: ASERTIVO {I strongly condemn the current terrorist attacks against Israel. I express my full solidarity with the victims, their families and loved ones.} **(4) Narendra Modi** (@narendramodi), ACTO DE HABLA: EXPRESIVO {Deeply shocked by the news of terrorist attacks in Israel. Our thoughts and prayers are with the innocent victims and their families. We stand in solidarity with Israel at this difficult hour.} **(5) Ursula von der Leyen** (@vonderleyen), ACTO DE HABLA: DECLARATIVO {Today, Hamas terrorists have struck at the heart of Israel capturing and killing innocent women and children. Israel has the right to defend itself - today and in the days to come. The European Union stands with Israel.}
La declaración de guerra del Estado de Israel	**(6) Israel לארשי** (@Israel), ACTO DE HABLA: «ADVISING» {We are at war. We will protect our citizens. We will not give in to terror. We will make sure that those who harm innocents pay a heavy price.}
El derecho a la defensa del Estado de Israel	**(7) Ursula von der Leyen** (@vonderleyen), ACTO DE HABLA: DECLARATIVO {I unequivocally condemn the attack carried out by Hamas terrorists against Israel. It is terrorism in its most despicable form. Israel has the right to defend itself against such heinous attacks.} **(8) Rishi Sunak** (@RishiSunak), ACTO DE HABLA: ASERTIVO {I am shocked by this morning's attacks by Hamas terrorists against Israeli citizens. Israel has an absolute right to defend itself. We're in contact with Israeli authorities, and British nationals in Israel should follow travel advice.} **(9) Joe Biden** (@POTUS), ACTO DE HABLA: DECLARATIVO {The world is seeing appalling images. Thousands of rockets raining down on Israeli cities. Hamas terrorists killing not only Israeli soldiers, but civilians on the streets and in their homes. It's unconscionable. Israel has a right to defend itself – full stop.}

Apoyo al Estado de Israel	**(10) Joe Biden** (@POTUS), ACTO DE HABLA: COMISIVO {Today, I spoke with @IsraeliPM about the appalling Hamas terrorist attacks in Israel. I offered our support and reiterated my unwavering commitment to Israel's security. @FLOTUS and I express our heartfelt condolences to the families who have lost loved ones.}
Rusia tiene responsabilidad sobre los acontecimientos	**(11) Thom Hartmann** (@Thom_Hartmann), ACTO DE HABLA: «ACCUSING» Hamas apparently knew how to get around Israel's Iron Dome defenses. They probably learned this from Iran. Iran almost certainly got the information from Russia. And who gave it to Russia? Sure looks like it was Donald Trump, at the request of Putin:
El ataque contra el Estado de Israel también representa un ataque contra America	**(12) Nikki Haley** (@NikkiHaley), ACTO DE HABLA: ASERTIVO {This is not just an attack on Israel—this was an attack on America. Finish them, @Netanyahu. They should have hell to pay for what they have just done.}
Los «luchadores» palestinos respectan los espacios sagrados, pero el Estado de Israel no	**(13) Keith Woods** (@KeithWoodsYT), ACTO DE HABLA: DECLARATIVO {While Palestinian fighters were given strict orders to respect churches and monasteries, Israel responded to the attacks by bombing a Mosque in Gaza.}

Tabla 1. Actos de habla y múltiples posiciones correspondientes a cada participante en el polílogo *online* sobre el conflicto Israel-Palestina (7 y 8 de octubre de 2023).

¿Qué sucede cuando conectamos y cruzamos estas diferentes posiciones? Siguiendo el modelo de diseño propuesto por Aakhus y Lewiński (2017 y 2023),[20] la figura 2 a continuación muestra las MPo como *clusters* para los AH. Los textos dentro de cada círculo representan las posiciones, mientras que los textos fuera de los círculos son los AH emitidos por los participantes (ya identificados en la tabla 1).

Es interesante observar, por ejemplo, que los AH (3), (4) y (5) utilizan diferentes fuerzas ilocutivas para expresar el mismo contenido proposicional de «rechazo al acto terrorista de Hamás». Estas manifestaciones de figuras de autoridad son importantes porque comunican información al público interesado en el tema del conflicto. Walton (2007) explica que, aunque este proceso de búsqueda y transmisión de información es un tipo de conversación teóricamente ignorado, forma parte de la deliberación inteligente porque permite captar información relevante sobre determinada situación (p. 62).

En la figura 2 a continuación se observa que los participantes están insertados en tensiones posicionales (Lewiński y Aakhus, 2023, p. 94). Estas MPo no representan posiciones

20 Me he inspirado totalmente en el trabajo de Aakhus y Lewiński (2017 y 2023) para diseñar estas interacciones.

que pueden reducirse diádicamente, es decir, configurarse como proposiciones en contra o a favor de un tema, sino que presentan un comportamiento no lineal.

Dado que la realidad polilógica permite una comunicación multinivel, la «declaración de guerra del Estado de Israel» se conecta a la posición de que «Rusia tiene responsabilidad sobre los acontecimientos». Estas dos posiciones establecen una falsa relación causal entre un hecho y otro y sugieren incorrectamente que la guerra implicaría la participación de Rusia. Además, la posición de que «Rusia tiene responsabilidad sobre los acontecimientos» se conecta con la posición de Biden que advierte «la intolerancia para aquellos que quieren sacar ventaja de la situación de Israel». El AH asociado a esta posición tiene la fuerza ilocutiva de acusación. Para Walton (2007), los AH de acusación públicos e indirectos revelan que un hablante tiende comprometerse emocional e ideológicamente con sus creencias.

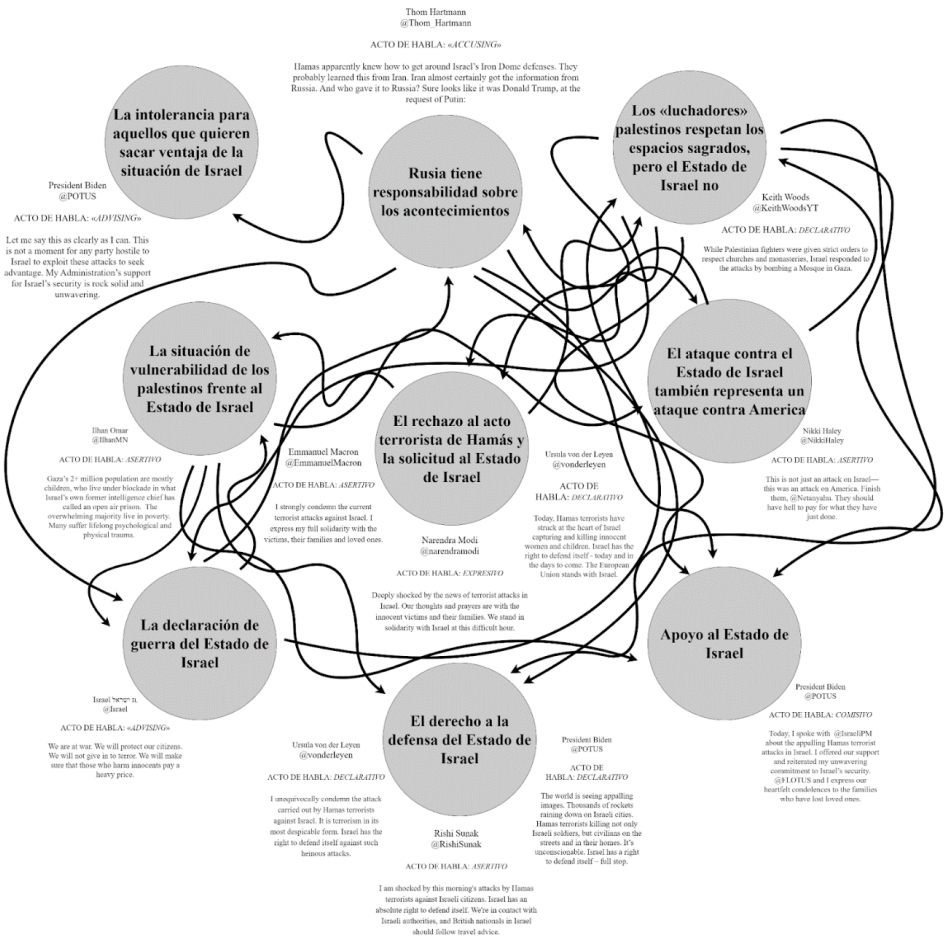

Figura 2. Representación del efecto de la cacofonía y el UG en los polílogos sobre el conflicto Israel-Palestina (2023) en *X*.

Dentro del modelo tradicional de argumentación deliberativa, la etapa de búsqueda de información es importante para persuadir los participantes a toma de decisiones. A diferencia del modelo tradicional, estas múltiples posiciones no se configuran como elementos persuasivos. En cambio, generan una cadena de otras posiciones, lo que refleja una estructura más compleja y menos orientada a la persuasión directa, como se mostrará en el ejemplo 2 a continuación.

4.2. Ejemplo 2

El ejemplo 2 destaca el AH público de Ursula von del Leyen (que corresponde al AH (7) en la tabla 1). Este AH puede ser organizado dentro de la posición sobre «el derecho de respuesta de Israel» y como se puede observar (figura 3), alcanzó más de 3.527 respuestas.

PARTICIPANTE: Ursula von der Leyen (@vonderleyen), presidente de la Comisión Europea (2019 — actualidad).

POSICIÓN: El derecho a la defensa de Israel.

ACTO DE HABLA: declarativo

Figura 3. Posición de Ursula von der Leyen acerca del ataque terrorista de Hamás contra Israel en *X*[21].

Es notable cómo una única publicación, configurada por un AH que expresa una posición específica, desencadena respuestas que generan cientos de nuevas posiciones y nuevos AH. Esto revela que la realidad polilógica se caracteriza por la coexistencia de múltiples posiciones que no necesariamente buscan persuadir, sino que se enfocan en presentar y verificar información diversa. Si se organizan, estos nuevos AH y sus fuerzas ilocutivas se agrupan en nuevos *clusters*, como se puede ver en la tabla 2 a continuación.

21 Enlace a la publicación original aquí, así como el enlace a la captura de la publicación en el sitio Webarchive.

Posiciones	Actos de habla
El apoyo a Israel representa el abandono del apoyo a Ucrania.	(1) *Accusing*: {Leaving project Ukraine too soon for a new one in Middle East?}
El apoyo de Europa a Israel refuerza el apartheid y el racismo.	(2) *Accusing*: {Germany stands on the side of apartheid as always. Notice how this hypocrite is happy when Ukrainians resist & fire missiles but not Palestinians. This is racism}
Una perspectiva negativa de Europa.	(3) *Accusing*: {Your evil empire will fall too}
La negación del Estado de Israel.	(4) *Assertive*: {Palestine is old 5000 years old country. Israel don't exists.} (5) *Assertive* {Free Palestine There is no country name Israel}
La legitimación del derecho a la defensa de Israel representa una traición a Europa.	(6) *Accusing*: {Si los estáis trayendo a Europa por cientos de miles. Abrid los ojos.}
Europa contribuye a la financiación de Hamás.	(7) *Directive:* {Then make sure that from now on no more financial means from Germany and the EU go to Hamas!} (8) *Interrogative*: {Thank you! Will the EU reconsider its (indirect) financial support of terrorist organizations such as Hamas and the PIJ (paid by European taxpayers) AND will the EU take appropriate diplomatic steps towards the Iranian regime, which is behind these attacks?}
La lógica del derecho de defensa de Israel justifica el derecho de defensa de Rusia contra Ucrania.	(9) *Assertive*: {Yeah, by that logic Russia has the right to defend itself against Ukraine.} (10) *Interrogative*: {Hmmm.. so you are accepting the terrorism which the Donbas regions endured from 2014. Russia had every right to go in and defend them?} (11) *Accusing*: {Two different stands..1.supporting Ukraine and against Russia 2.Supporting Israel and against Palastine.. Hypocrisy...} (12) *Interrogative*: {And Donbass?} (13) *Assertive*: {Israel has the rights, but Russia shouldn't when it comes to Ukrainazis.}

Ambos (Israel y Palestina) tienen derecho a la auto defensa.	(14) *Assertive*:{I believe that this principle applies universally: if Israel has the right to defend itself, so does Palestine. Israel should cease taking their land and return to the pre-1967 borders; everything will be resolved as simple as that.} (15) *Assertive*: {Indeed, Palestine has the right to defend itself against the zionist.} (16) *Expressive*: {What about Palestine? They are the persecuted, defending themselves as best they can. Dismantle Illegal Israel if you want peace.} (17) *Interrogative*: {Does Palestine have the right to defend themselves when their homes were bombed by Israel in May?} (18) *Assertive*: {Palestinians are not terrorists; they are defending their homeland.} (19) *Assertive*: {So do Palestinians has very right to defend them self from invaders and settlers} (20) *Assertive*: {Israel has a right to defend itself, but Palestine does not. This is the height of hypocrisy.}
En defensa de la libertad de Palestina.	(21) *Exhortación*: {Free Palestine} (22) *Assertive*: {Palestinian's have the right to defend and safeguard their existence.} (23) *Assertive*: {It's the opposite, Palestinians are defending themselves... Free Palestine.} (24) *Exhortación*: {All lives matter #FreePalestine} (25) *Exhortación*: {We stand with P@léstiné} (26) *Assertive*: {Fighting back against those things = Universal condemnation. Ethnic cleaning and apartheid = Fine.} (27) *Accusing*: {You're an unelected corrupt politician, no one cares what you have to say, just like you didn't care when Israeli terrorists were killing innocent civilians in Palestine. Palestine will be FREE} (28) *Exhortación*: {Palestine Will be free they're fighting for their freedom!}

El ejército de Israel es terrorista.	(29) *Assertive*: {Terrorism is a pejorative. The Israeli army are a textbook example of it. But you refuse to call IDF terrorists - Palestine has the right to defend itself against heinous attacks by Zionists ethnically cleanse land because think they're God's special people} (30) *Advising*: {You don't have eyes, but still we show you what is the truth.I think we will see now who is a terrorist. ?}
Critica a la parcialidad de Europa a favor de Israel.	(31) *Assertive*: {The world's response; Silence during daily Israeli bombings in #Palestine #Gaza, but now, with Hamas' attacks on Israel, there's attention. This is an absolute case of hypocrisy from the world leaders.} (32) *Accusing*: {Terrorism that Palestinians had to endure for centuries, but you were always quiet about that you hypocrite} (33) *Accusing*: {EU the true face of hypocrisy} (34) *Accusing*: {With that logic it is Israel who is the oppressor like Russia and been grabbing land inch by inch over the years. All these years the EU nor UK or the US condemned those acts, instead funded the settlements. Double standards Ursula!} (35) *Expressive*: {What a surprise! Missing when Israel bombs civilians} (36) *Accusing*: {hypocrisy at its finest form}

Tabla 2. Las múltiples posiciones como *clusters* para los actos de habla en respuesta a la publicación de von der Leyen sobre el conflicto Israel-Palestina (octubre de 2023).

La figura 5, a continuación, también muestra visualmente cómo las MPo organizan los AH, operando como *clusters*, y específicamente la frecuencia con la que cada tipo de actos aparece en los *clusters*. Aquí, el análisis se centra exclusivamente en las respuestas más destacadas, omitiendo la amplia cadena de interacción, es decir, las respuestas a las respuestas y así sucesivamente.

Como mencioné anteriormente, la diversidad de intenciones y creencias contenidas en los AH permite organizarlos en *clusters* proposicionales. Esto posibilita una estructura de análisis para identificar qué tipos de sesgos cognitivos y falacias interfieren en la argumentación. Por ejemplo, la afirmación de que «el apoyo a Israel representa el abandono del apoyo a Ucrania» presenta un razonamiento dicotómico. Del mismo modo, la posición que sostiene que «el apoyo de Europa a Israel refuerza el apartheid y el racismo» se basa en un razonamiento excesivamente generalizado. Además, la posición que afirma que «ambos

(Israel y Palestina) tienen derecho a la autodefensa» revela un sesgo de confirmación al presuponer automáticamente ambas partes tienen igual y justificado derecho a la autodefensa, sin considerar las responsabilidades involucradas. Estas posiciones mayoritariamente asertivas indican que los hablantes expresan presunciones de verdad arraigadas. Identificar estos patrones a gran escala puede contribuir significativamente a la orientación del debate.

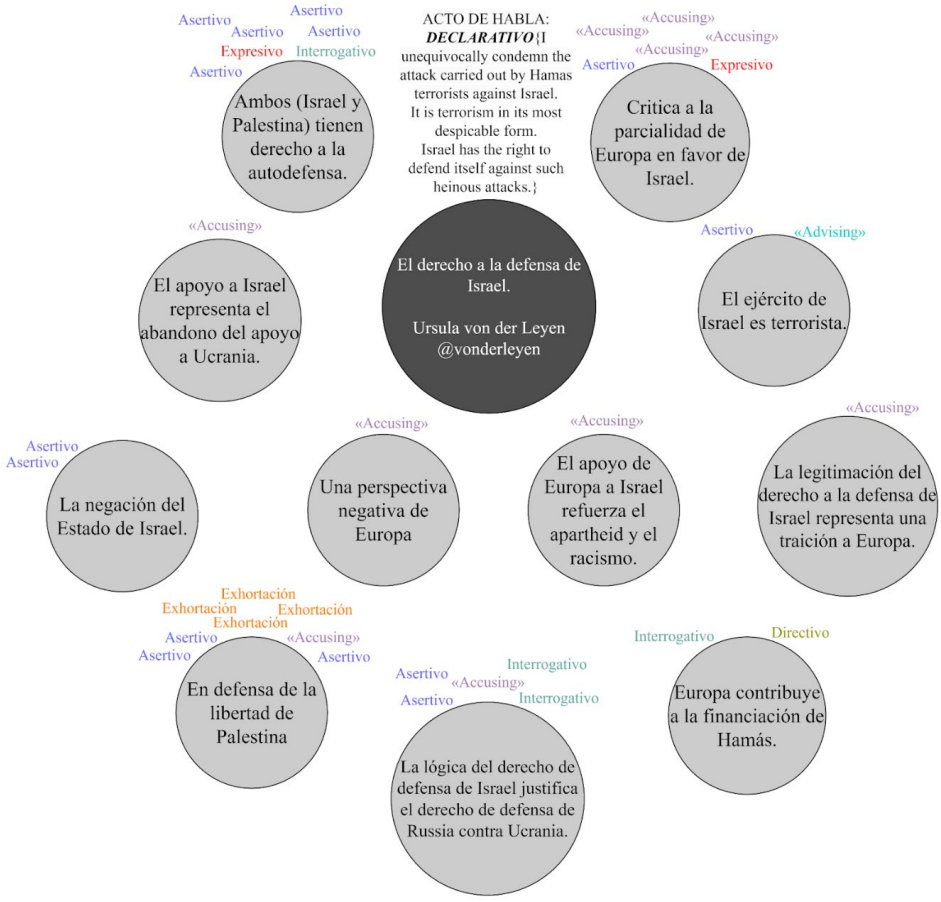

Figura 4. Las múltiples posiciones y actos de habla en respuesta a Ursula von der Leyen sobre el conflicto Israel-Palestina (7 de octubre de 2023)

Como se puede observar en la figura 4, algunas posiciones generan más actos de habla que otras. Para entender este fenómeno, Sylvie Bruxelles y Catherine Kerbrat-Orrecchioni (2002) proponen que los polílogos permiten la formación de coaliciones. Según ellas, las coaliciones en los polílogos se refieren a alianzas temporales entre distintos grupos de participantes que se unen para apoyar mutuamente sus argumentos y objetivos comunes. Esto se puede observar a través de: (i) los marcadores de acuerdo, como palabras «sí», «exactamente», «correcto»,

etc.; (ii) apoyo léxico, por ejemplo, el aporte de los participantes con información para complementar expresiones vagas, incompletas, etc.; (iii) ayuda mutua para fortalecer la fuerza ilocutiva y alcanzar los objetivos argumentativos y, por último, (iv) el uso de los pronombres personales por parte de los participantes, por ejemplo, «nosotros», «usted», etc. (p.76-82).

En la figura 5 se ilustra cómo se forma la coalición a través del uso de pronombres personales como marcadores. Para facilitar el análisis, el término «palestino» puede ser equiparado con el pronombre personal «él» y palestinos con «ellos». Además, se han resaltado en rojo las líneas y posibles pronombres para mejorar la visualización.

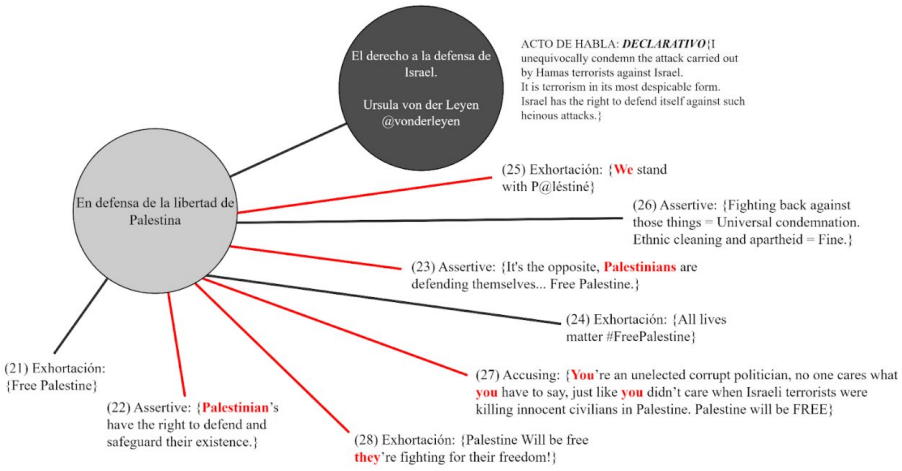

Figura 5. El uso de los pronombres personales como marcadores de coalición en la posición que defiende la libertad de Palestina.

En la publicación de von der Leyen, se observa la predominancia del AH asertivo, seguido por el acto de acusación, interrogación, exhortación, expresión y directiva (Tabla 3). Según Kauffeeld e Innocenti (2018), la intención del AH asertivo es afirmar creencias como verdaderas.

Tipo de acto de habla	Número de veces que aparece en la interacción *online*
Asertivo	14
«Accusing»	10
Interrogativo	4
Exhortación	4
Expresivo	2
Directivo	1

Tabla 3. El número de veces que cada acto de habla aparece como respuesta a la publicación de von der Leyen

4.3. Ejemplo 3

El ejemplo 3 identifica una de las publicaciones de Biden (figura 6) sobre el conflicto Israel-Palestina. La posición, en este caso, se expresa por un AH que aconseja sobre «la intolerancia hacia quienes se aprovechan de la situación de Israel».

PARTICIPANTE: Joe Biden (@vPOTUS), presidente de USA.
POSICIÓN: La intolerancia hacia quienes se aprovechan de la situación de Israel.
ACTO DE HABLA: «*Advising*»

Figura 6. La posición de Joe Biden acerca del ataque terrorista de Hamás contra Israel en la red social *X*. [22]

Como se puede observar en la figura 7, la publicación generó posiciones diferentes entre las respuestas más relevantes[23]. Las posiciones en respuesta muestran la pluralidad de AH entre los participantes.

22 Enlace para la publicación original aquí y también el enlace para la captura de la publicación en el site Webarchive.

23 Dentro de la metodología aplicada.

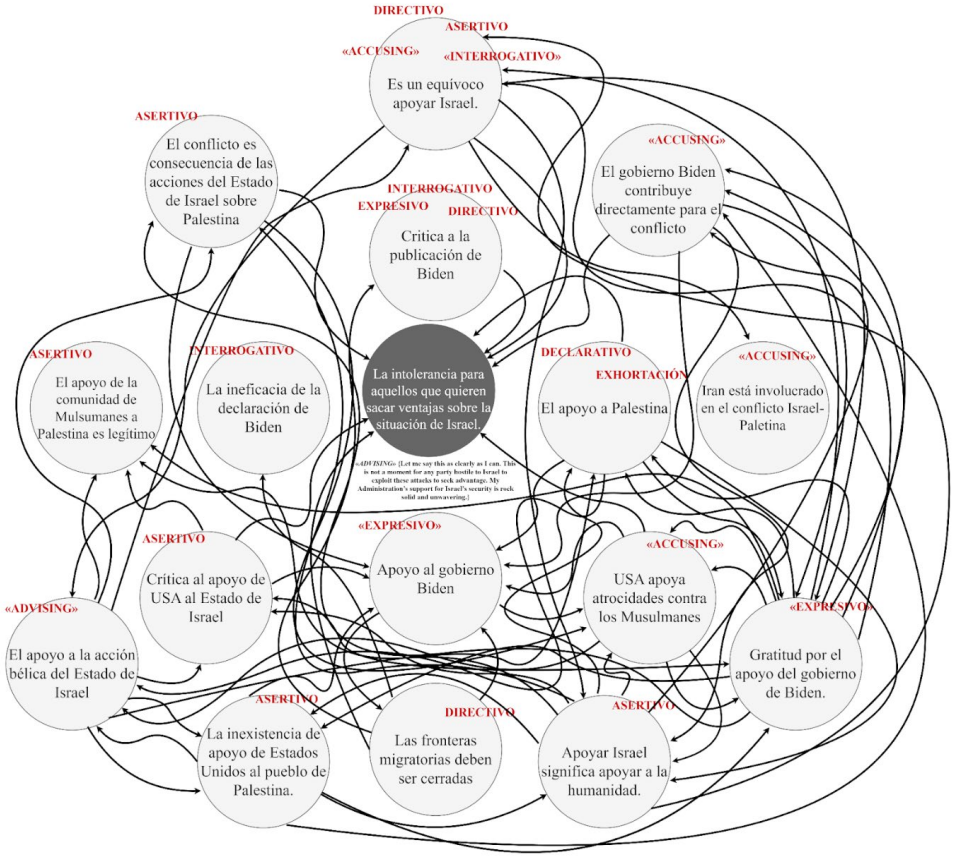

Figura 7. Las múltiples posiciones y los tipos de actos de habla en respuesta a la publicación de Joe Biden en *X*

Según Walton (2007), la fuerza ilocutiva de un AH indica el tipo de compromiso de un hablante con una proposición. Es decir, señala qué tipo de compromiso está implicado dentro de la dinámica comunicativa. Es más, explica Walton, indican cómo argumentó y qué posiciones tomó. Aunque nuestras creencias y nuestros compromisos tengan una relación directa, no es necesario un conocimiento profundo de las creencias reales de un argumentador para evaluar su argumento. Si, como afirma Walton, el compromiso personifica la creencia, entonces, identificar los tipos de compromisos presentes en la dinámica comunicativa puede contribuir a entender lo que está en juego en la comunicación. En la tabla 4 a continuación se observa que el acto de habla de acusación (*accusing*) predomina, seguido por el acto de habla asertivo, interrogativo, expresivo, directivo, de «*advising*», declarativo y exhortativo. El compromiso implícito en un AH de acusación, por ejemplo, es destacar una acción inadecuada del acusado, una creencia que el acusador tiene de que el acusado ya sabe lo que se le está alegando. Al acusar, se da a entender que lo que el acusado hizo es

censurable o reprensible (Kauffeld y Goodwin, 2022). Además, Walton (2007) explica que AH como de acusación, cuando se emiten en serie, revelan fanatismo y demuestran que los hablantes están cerrados para la razón porque demuestran sesgos endurecidos.

Tipo de acto de habla	Número de veces que aparece en la interacción *online*
«Accusing»	19
Asertivo	7
Interrogativo	4
Expresivo	4
Directivo	3
«Advising»	1
Declarativo	1
Exhortación	1

Tabla 4. El número de veces que cada acto de habla aparece como respuesta a la publicación de Biden.

En condiciones normativas, un acto de habla de acusación, según Kauffeld y Goodwin (2022), implica una obligación de respuesta del acusado hacia el acusador. Sin embargo, en los polílogos *online*, este requisito no se cumple porque, tal como explica Walton (2007), cuando los hablantes están involucrados en un comportamiento que explicita fanatismo, están cerrados a la razón y presentan sesgos cognitivos que impiden el desarrollo de la conversación y la consideración de diferentes perspectivas.

Considero que los ejemplos citados captan una dinámica general observada en las redes sociales, específicamente en *X*, y tienen como objetivo ilustrar patrones recurrentes en la argumentación *online* y cómo se manifiestan en la pluralidad de AH.

5. Conclusiones

El estudio de cómo los colectivos argumentan actualmente en el contexto de desacuerdos complejos en las conversaciones *online*, se hace cada vez más necesario. En este trabajo propuse que existe un *uncomoon ground* (UG) fijado en el caso de un polílogo argumentativo *online*. Este desacuerdo presente en los polílogos *online* emerge del UG entre los participantes y puede ser analizado desde la pluralidad de los actos de habla (AH) presentes en las múltiples posiciones (MPo). Si eso es así, entonces, tal como plantean Aakhus y Lewiński (2023), el polílogo puede ser considerado una herramienta que permite establecer estrategias para la resolución racional de nuestros desacuerdos. En los polílogos *online*, la pluralidad de AH es amplificada por el efecto de larga escala de las redes sociales. Por lo tanto, la perspectiva tradicional de la pragma-dialéctica no es suficiente para analizar conversaciones *online*. En cambio, parte del desafío presente en las teorías de la argumentación y en las teorías de los actos de habla es identificar y explicar los patrones presentes en los polílogos.

Este trabajo contribuye a afrontar el anterior desafío. Al analizar diferentes ejemplos de decisiones en *X*, no me centré tanto en la cualidad de lo dicho en los ejemplos, sino en la pluralidad de lo dicho dentro de un polílogo, evidenciando el pluralismo ilocutivo. Demostrar la relación entre la falta de *common ground* y la pluralidad de AH no es fácil, pero los ejemplos presentados permitieron identificar que, al menos en lo que concierne a la red social *X*, las MPO actúan como *clusters* para los múltiples AH.

En cada uno de los *clusters* posicionales, la ocurrencia de tipos de AH varía. En algunos *clusters*, las fuerzas ilocutivas se repiten, mientras que en otros presentan una mayor diversidad de actos de habla. Pero, en general, la pluralidad de AH se confirma, indicando así que diferentes fuerzas ilocutivas se alternan para introducir nuevas informaciones en la conversación.

Aunque los ejemplos presentados sean insuficientes para generalizar la identificación de patrones definitivos, cabe destacar una mayor incidencia de AH asertivos y de acusación. Este resultado exploratorio está abierto a análisis futuros más sistemáticos. En los ejemplos presentados, se pueden observar posiciones contradictorias, complementarias, y también posiciones que desvían del tema en cuestión o que aprovechan determinadas publicaciones para hacer asociaciones aleatorias. Esto indica que, en los polílogos, los participantes no cumplen con las condiciones normativas ni con la función epistémica de reconocimiento de las intenciones del hablante.

Un ejemplo de ello son los diversos AH de acusación dirigidos a la publicación de Biden o la asociación fuera de contexto entre el conflicto Israel-Palestina y Rusia-Ucrania en la publicación de von der Leyen. Los ejemplos también permitieron identificar que, tal como plantearon Bruxelles y Kerbrat-Orrecchioni (2004), los polílogos permiten la formación de coaliciones entre los múltiples participantes desde los múltiples lugares.

Referencias bibliográficas

Aakhus, M., & Lewiński, M. (2017). Advancing Polylogical Analysis of Large-Scale Argumentation: Disagreement Management in the Fracking Controversy. *Argumentation*, *31*(1), 179-207.

Arielli, E. (2018). Sharing as Speech Act. *Versus*, *127*, 243-258.

Battersby, M., & Bailin, S. (2011). Critical Inquiry: Considering the Context. *Argumentation*, *25*(2), 243-253. https://doi.org/10.1007/s10503-011-9205-z

Baym, N. K. (2015). *Personal Connections in the Digital Age*. John Wiley & Sons.

Borg, E., & Connolly, P. J. (2022). Exploring Linguistic Liability. En E. Lepore & D. Sosa (Eds.), *Oxford Studies in Philosophy of Language Volume 2*. Oxford University Press.

Bruxelles, S., & Kerbrat-Orecchioni, C. (2004). Coalitions in Polylogues. *Journal of Pragmatics*, *36*(1), 75-113.

Camp, E. (2017). Pragmatic Force in Semantic Context. *Philosophical Studies*, *174*(6), 1617-1627.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Corredor, C. (2020). Speaking, Inferring, Arguing. On the Argumentative Character of Speech. *Studia Semiotyczne*, *34*(2), 43-64.

Danaher, J. (2024). How Technology Alters Morality and Why It Matters [Ethics]. *IEEE Robotics & Automation Magazine*, *31*(2), 147-148. https://doi.org/10.1109/MRA.2024.3388278

Dennett, D. C. (2004). *La evolución de la Libertad*. Grupo Planeta (GBS).

Eemeren, F. H. van, & Grootendorst, R. (2003). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616389

Eemeren, F.H. y Houtlooser (2007). Countering Fallacious Moves. *Argumentation*, 21, 243-252.

Frápolli, M. J. (2024). *Russell's Paradox* [Webnar]. Seminario Permanente PHIDELO https://www.youtube.com/watch?v=valWvPrL0EY

Greco, S. (2023). Twitter Activists' Argumentation Through Subdiscussions: Theory, Method and Illustration of the Controversy Surrounding Sustainable Fashion. *Argumentation*, *37*(1), 1-23.

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*(3), 377-388.

Innocenti, B. (2022). Demanding a Halt to Metadiscussions. *Argumentation*, *36*(3), 345-364.

Kauffeld, F. J., & Goodwin, J. (2022). Two Views of Speech Acts: Analysis and Implications for Argumentation Theory. *Languages*, *7*(2), Article 2. https://doi.org/10.3390/languages7020093

Kauffeld, F. J., & Innocenti, B. (2018). A Normative Pragmatic Theory of Exhorting. *Argumentation*, *32*(4), 463-483. https://doi.org/10.1007/s10503-018-9465-y

Kerbrat-Orecchioni, C. (2004). Introducing Polylogue. *Journal of Pragmatics*, *36*(1), 1-24. https://doi.org/10.1016/S0378-2166(03)00034-1

Landemore, H. (2021). 2. Open Democracy and Digital Technologies. En 2. Open Democracy and Digital Technologies (pp. 62-89). University of Chicago Press. https://doi.org/10.7208/9780226748603-003

Lewinski, M. (2021). Speech Act Pluralism in Argumentative Polylogues. *Informal Logic*, *42*(4), 421-451.

Lewiński, M., & Aakhus, M. (2014). Argumentative Polylogues in a Dialectical Framework: A Methodological Inquiry. *Argumentation*, *28*(2), 161-185.

Lewiński, M., & Aakhus, M. (2023). *Argumentation in Complex Communication: Managing Disagreement in a Polylogue*. Cambridge University Press.

Lewiński, M., Cepollaro, B., Oswald, S., & Witek, M. (2023). Norms of Public Argument: A Speech Act Perspective. *Topoi*, *42*(2), 349-356.

Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, *8*(1), 339-359. https://doi.org/10.1007/BF00258436

Macagno, F., & Capone, A. (2016). Uncommon Ground. *Intercultural Pragmatics*, *2*(13), 151-180.

Marcoccia, M. (2004). On-line polylogues: Conversation structure and Participation Framework in Internet Newsgroups. *Journal of Pragmatics*, *36*(1), 115-145. https://doi.org/10.1016/S0378-2166(03)00038-9.

Marsili, N. (2021). Lies, Common Ground and Performative Utterances. *Erkenntnis*, *88*(2), 567-578.

Mclver Lopes, D. (2014). Aesthetic Appreciation. En *Beyond Art*. Oxford University Press.

Meijers, A. (2007). Collective Speech Acts. En *Intentional Acts and Institutional Facts* (Vol. 41, pp. 93-110). Springer, Dordrecht.

Musi, E., & Aakhus, M. (2018). Discovering Argumentative Patterns in Energy Polylogues: A Macroscope for Argument Mining. *Argumentation*, *32*(3), 397-430.

Palmieri, R., & Mazzali-Lurati, S. (2016). Multiple Audiences as Text Stakeholders: A Conceptual Framework for Analyzing Complex Rhetorical Situations. *Argumentation*, *30* (4), 467-499.

Prigogine, I., & Stengers, I. (2018). *Order Out of Chaos: Man's New Dialogue with Nature*. Verso Books.

Ramos Mendonça, B. (2012). Conhecimento Simbólico na Álgebra da Lógica de Venn. *Principia: an international journal of epistemology*, *16*(3), 471-488. https://doi.org/10.5007/1808-1711.2012v16n3p471

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5-6), 701-721.

Tversky, Amos & Kahneman, Daniel (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185 (4157).1124-1131.

Walton, D. N. (2008). *Argumentation schemes*. Cambridge ; New York : Cambridge University Press.

Walton, D. (2007). *Media Argumentation: Dialectic, Persuasion and Rhetoric*. Cambridge University Press. https://doi.org/10.1017/CBO9780511619311

Williamson, T. (2002). *Knowledge and Its Limits*. Oxford University Press.

Wolff, R. (2019). Towards a Critical Theory of the Technosystem. *Jus Cogens*, *1*(2), 173-185. https://doi.org/10.1007/s42439-019-00011-z

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 119-135
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.611821

¿Es la inteligencia artificial doxástica un igual epistémico?*

Is doxastic artificial intelligence an epistemic peer?

*ALBERTO MURCIA CARBONELL***

Resumen: La inteligencia artificial doxástica (IAD) es un tipo de inteligencia artificial que reproduce actitudes doxásticas. Si la IAD cumple con las condiciones de paridad epistémica que se le exige a un humano, ¿podría ser también un igual epistémico? Dos iguales epistémicos sostienen propiedades cognitivas simétricas como la inteligencia, el razonamiento o la ausencia de sesgos. Para evaluar si alguien es un igual se tendrán en cuenta estas condiciones: (1) la igualdad probatoria, (2) la igualdad cognitiva y (3) revelación completa. La IAD cumple tanto (1) como (2), pero es en (3) cuando se descubre que no puede ser un igual epistémico. Ésta responde con la opinión popular más aceptada estadísticamente, es incapaz de sostener y defender sus propias afirmaciones y éstas no son un genuino acto de habla. Es una máquina doxástica incapaz de señalar cuáles son las razones que guían su respuesta.
Palabras clave: desacuerdos epistémicos, epistemología, cultura digital, autoridad epistémica, inteligencia artificial, entrenamiento automatizado.

Abstract: Doxastic artificial intelligence (DAI) is a subclass of artificial intelligence that reproduces doxastic attitudes. If DAI meets the conditions for epistemic peerhood that are required for a human, could it also be considered as an epistemic peer? Two epistemic peers hold symmetric cognitive properties such as intelligence, reasoning, or absence of bias. To assess whether someone is a peer, these conditions will be taken into account: (1) evidential equality, (2) cognitive equality, and (3) situation of full disclosure. DAI meets both (1) and (2), but it is (3) when it is revealed that it cannot be an epistemic peer. Its answers are the most accepted popular opinion statistically, it is unable to sustain and defend its own asseverations, which are not considered as a genuine speech act. It is a doxastic machine unable to point out what are the reasons guiding its answers.
Keywords: epistemic disagreements, epistemology, digital culture, epistemic authority, artificial intelligence, machine learning.

Recibido: 11/04/2024. Aceptado: 24/04/2024.

* Este artículo fue posible gracias al proyecto nacional de investigación *Meta-actitudes, desacuerdos profundos y progreso moral*, financiado por el Ministerio de Ciencia e Innovación, PID2021-124152NB-I00. Universidad Carlos III de Madrid.

** Universidad Carlos III de Madrid. Profesor Asociado. Las líneas de investigación principales son: 1) filosofía y tecnología, 2) ética y nuevos medios, 3) cultura digital y 4) *game studies*. Última publicación: Murcia, A. (2022). La justificación del victimario como estrategia identitaria para el veterano de guerra. *Dilemata*, (39), 61-78. Correo electrónico: amurcia@uc3m.es

1. Introducción

La inteligencia artificial (IA) reproduce comportamiento inteligente bastante sofisticado. Puede auditar las cuentas de una compañía, determinar el reparto de ayudas sociales, reconocer el rostro de una persona en diferentes contextos, dirigir un vehículo u ordenar el inventario de un almacén para mejorar la eficiencia logística, por mencionar algunos ejemplos. Incluso es capaz de reproducir actitudes doxásticas, como ofrecer un diagnóstico médico, escribir un ensayo sobre un tema, responder a dilemas morales, trazar la mejor ruta para llegar hasta un destino determinado, o incluso discutir sobre temas complejos de calado filosófico. Dado que muestra estas propiedades doxásticas que parecen de fiar, ¿podríamos llegar a considerar que una IA es un igual epistémico?

Antes de atajar la cuestión, consideraremos que la IA pertenece al grupo de la tecnología inteligente [*smart technology*][1]. La tecnología inteligente son artefactos cuya función es reproducir comportamiento inteligente (Floridi 2023). No todas las tecnologías inteligentes son IA, pero la IA siempre es una tecnología inteligente. Un lavaplatos, una máquina automática de planchado, o un robot aspiradora, reproducen comportamiento inteligente (lavar platos, planchar, barrer), pero no son IA. De entre las tecnologías inteligentes que son IA, tendremos en cuenta solo las que reproducen un tipo de comportamiento: las actitudes doxásticas. Todas las IA reproducen comportamiento inteligente, pero no todas reproducen actitudes doxásticas. Una IA de reconocimiento facial no reproduce actitudes doxásticas, Dall-e tampoco. Un navegador inteligente, que traza una ruta entre dos espacios geográficos calculando el mejor tiempo de llegada, sí que reproduce actitudes doxásticas. ChatGPT, como generador de textos, también reproduce actitudes doxásticas. Llamaremos inteligencia artificial doxástica (IAD) a este tipo de IA, y será el objeto de este estudio.

Para considerar si una IAD puede ser un igual epistémico primero hay que entender qué es un desacuerdo epistémico. Éstos se producen cuando dos personas, que se consideran iguales epistémicos, sostienen actitudes doxásticas diferentes después de haber examinado la misma evidencia de carácter público (Frances y Matheson 2019, Zagzebski 2012).[2] Una de las consecuencias de estar en desacuerdo es que cuando se nos revela que alguien a quien consideramos como un igual epistémico sostiene que no-P, podría ser razonable que dejemos de confiar en nuestra creencia sobre P (Cocchiaro y Frances 2021, p. 1063, Kelly 2013). Los iguales comparten simetría en sus capacidades cognitivas, en aquello que saben sobre P y en el acceso a las evidencias. A la opinión de un igual siempre se le otorga cierto crédito, ya que se le considera como autoridad epistémica, aunque sea una autoridad de tipo débil, esto es, la autoridad del experto que lo es únicamente en el ámbito de las proposiciones sobre P (Zagzebski 2012).[3] Si alguien se ve en la posible obligación de re-evaluar su creencia

1 Traducimos "smart technology" por "tecnología inteligente", aunque consideramos que el término es confuso, pues contribuye a la idea errónea de que son "inteligentes". En términos de Floridi (2023), son *listas, ingeniosas o sagaces*, pero no son inteligentes. Sucede lo mismo con "inteligencia artificial", que da a entender algo incorrecto sobre esta tecnología, esto es, que produce inteligencia.

2 En concreto, aquí nos referimos a desacuerdos razonables, "casos en los que personas inteligentes con acceso a la información disponible y relevante llegan a conclusiones incompatibles" (Felman 2011, p.416).

3 Una autoridad epistémica no siempre es un igual epistémico, pero un igual epistémico siempre es una autoridad epistémica desde el punto de vista del agente que evalúa la paridad. En otras palabras, para que A considere

es porque tiene la certeza de que la otra persona es un igual epistémico. Así, antes de que llegue el desacuerdo, primero tendrá que determinar si el otro es un igual, y no a la inversa, es decir, que sea el desacuerdo el que le muestre que se encuentra ante un igual.

Para saber si se está ante un igual epistémico se constata si ese alguien cumple ciertas condiciones de simetría que permitan evaluarlo como tal. Alguien es un igual si posee ciertas virtudes epistémicas (Zagzebski 2012, p. 205), como la inteligencia, la capacidad para razonar y la ausencia de sesgos (Rowland 2017, Kelly 2005), a lo que hay que añadir que sus antecedentes garantizan que es conocedor de P (Christensen y Lackey 2013). El igual no se circunscribe solo a *alguien*, sino que también podría ser *algo*. Si una IAD cumple las condiciones de simetría, entonces se debe afirmar que también es un igual epistémico. Señalamos a continuación éstas condiciones, elaboradas desde Bistagnino (2011), pero que se siguen igualmente desde Kelly (2005 y 2013), Christensen (2009), Lackey (2013), Rowland (2017) Frances y Matheson (2019) y Cocchiaro y Frances (2021).

(1) *Igualdad probatoria*: A y B conocen las evidencias y los argumentos que influyen en la cuestión de P.

(2) *Igualdad cognitiva*: A y B son igualmente virtuosos en términos epistémicos en su evaluación de las pruebas y argumentos que influyen en P.

(3) *Revelación completa*: A y B comparten todas sus pruebas y argumentos sobre P.

Al considerar a alguien (o algo) como un igual, se evalúa su comportamiento como expresión de su cognición, prestando atención también al historial de eficiencia y competencia del agente sobre P (Christensen y Lackey 2013). Si la evaluación es positiva, entonces A tiene razones para confiar en B como su igual epistémico, el cual también es reconocido con cierta autoridad epistémica. Dicho esto, ¿podemos considerar a una IAD como un igual epistémico?

Supongamos lo siguiente. Javier es un taxista experimentado. Acaba de recoger a un cliente que quiere llegar lo antes posible al aeropuerto. ¿Cuál es la mejor ruta? Javier conoce bien las carreteras y está acostumbrado a las contingencias de su ciudad a esas horas. Tiene un navegador con IAD que, para tomar un ejemplo en concreto, diremos que es Google Maps. Javier puede seguir varias estrategias: (i) Conecta Maps y deja al dispositivo que sugiera la mejor ruta; (ii) Prefiere ser él quien decide la ruta y no lo conecta; (iii) Javier evalúa la ruta a la vez que activa Maps para que también le proporcione un recorrido hasta el aeropuerto.

Podemos pensar (i) de dos formas: o bien Javier está tomando Maps como un superior epistémico, o bien considera que es un tipo de autoridad epistémica débil. En el primer caso, Javier evalúa como más probable, digamos un 90%, que la IAD trace una ruta mejor de la que él pueda pensar. Javier asume que el dispositivo es virtuoso en tanto que eficiente. A eso se le llama "deferencia completa" (Elga 2007): ante una proposición sobre la ruta óptima de

---

a B como un igual epistémico sobre P, A está obligado a afirmar que B es una autoridad epistémica sobre P. Siguiendo a Zagzebski (2021, p.1), utilizamos el concepto de "autoridad epistémica" como "autoridad epistémica débil", es decir, alguien que es un "experto" sobre un tema, un área de conocimiento, una afición, etc. Una cuestión diferente es por qué el experto es autoridad. Se podrá observar que en el artículo se ha simplificado bastante la cuestión, asumiendo que el experto es alguien (o algo) que es capaz de producir enunciados verdaderos de forma estable y efectiva sobre un asunto determinado.

llegada al aeropuerto, las posibilidades de que Javier crea que P son las mismas que tiene Maps de considerar que P. En el segundo caso, podemos pensar en Maps como una autoridad epistémica débil. Siguiendo el sentido que le da Zagzebski (2012, p.3), primero debemos partir desde la premisa de que Javier tiene la intención de tener certeza; después, Javier puede considerar *X* como la mejor ruta, pero si Maps responde que *Y* tendrá un mejor tiempo estimado de llegada, Javier, que quiere estar en lo correcto, desechará su creencia sobre *X* y depositará su confianza en *Y*. Javier considera a la IAD como una autoridad epistémica, pero no como un igual epistémico.

En (ii), Javier cree saber sin duda cuál es la mejor ruta. Sus años conduciendo le dan la confianza necesaria como para justificar su creencia. Así que no utilizará Maps. En esta situación, Javier se considera como la única autoridad epistémica.

En (iii), la opinión de Javier sobre cuál es la ruta óptima no depende de lo que Maps sugiere, sino que va a comparar lo que él cree con lo que responda el navegador. Suele pasar que la ruta trazada por Maps coincide con la que Javier opina que es la óptima, por lo que en esta ocasión también confía en que será igual. De esta manera, Javier trata a Maps como si fuera un igual epistémico, pero también como una autoridad epistémica, porque suele ser efectiva en su función. Es (iii) la situación que nos interesa para este artículo. ¿Está Javier en un error al evaluar que la IAD puede ser un igual epistémico?

En lo que sigue no indagaremos sobre si podemos tener un desacuerdo razonable genuino con una IA, tampoco si el desacuerdo *debería* obligarnos a reconsiderar nuestras creencias, ni si en la revelación completa los iguales deben discutir sobre sus argumentos y razones. Queremos centrarnos únicamente en la pregunta de si una IAD puede ser considerada como un igual epistémico. Tal vez se pueda inferir de nuestras palabras que estamos asumiendo que se puede tener un desacuerdo con una IA, puesto que ser un igual es condición necesaria para el desacuerdo epistémico. No es así. La posición que se defenderá aquí es que la IAD no puede ser considerada como un igual epistémico, y si esto es condición necesaria para estar en un desacuerdo, entonces no podemos tener un desacuerdo. Sea como sea, como dijimos, no trataremos este asunto aquí.

Afirmamos que la IAD puede cumplir con las condiciones de (1) igualdad probatoria y de (2) igualdad cognitiva, pero (3) la revelación completa descubre que la naturaleza estadística de las opiniones reproducidas es insuficiente para que la consideremos como igual y como autoridad epistémica. Las razones sobre por qué no es un igual las daremos en la sección 3, mientras que en la 4 dedicaremos un breve espacio a explicar las razones por las que consideramos que tampoco es una autoridad epistémica. Antes de llegar a ese punto, primero revisaremos cuál es el valor doxástico de una IAD.

2. El valor doxástico de la IAD

La IAD es un tipo de artefacto que reproduce el comportamiento inteligente de las actitudes doxásticas. Una IAD como ChatGPT representa estos datos mediante lenguaje natural, mientras que un navegador inteligente lo hace señalando una ruta desde una posición geográfica hasta otra que tenga el mejor tiempo de llegada. Sea mediante palabras o con un trazado sobre un mapa, ambas tecnologías cumplen con su tarea presentando los datos

en forma de enunciados declarativos, veraces y con valor doxástico.[4] La IAD sustituye la inteligencia humana mediante procesos que son agenciales, pero no inteligentes (Floridi, 2023, pp. 23-24).

Es pertinente preguntarse si una IAD puede ser considerada como igual epistémico debido a las características señaladas, pero también por su irrupción en diversas actividades humanas, como en la medicina, la enseñanza o la auditoría de empresas, entre otras. Esta pregunta no se ha planteado como tal en las discusiones sobre el tema de la igualdad epistémica. Podemos señalar al menos un factor que consideramos que pudo haber influido en que el desdén hacia esta cuestión, que tiene que ver con las propiedades epistémicas de las computadoras. Se podría resumir en si una computadora, y por extensión una IA, *sabe* algo sobre aquello que computa. Desde el desafío de la "habitación china" que planteó John Searle (1982) se considera que las computadoras carecen de capacidades semánticas. El argumento de Searle, expuesto aquí de manera muy tosca, señala que las computadoras son máquinas que operan solo con sintaxis mediante reglas bien definidas en su programa. Son eficientes en su función de computar sin que sea necesario que "entiendan" el contenido semántico que transmiten sus operaciones. En otras palabras, manejan la sintáctica del lenguaje, pero no sus significados; producen información semántica, pero para que computen no hay necesidad de que sean máquinas semánticas.

Si seguimos la propuesta de Searle, esta nos lleva a descartar que una IA comparta las propiedades epistémicas de un humano. Al mismo tiempo, también podemos afirmar que, pese a esta limitación, pueden ser eficientes en su función. Que una IA desconozca el significado de la palabra "cáncer" no impide que señale un cáncer en una tomografía con un grado de acierto equivalente al de un oncólogo experimentado. Las computadoras no necesitan semántica para actuar y ser eficientes. Por lo tanto, la cuestión no sería la de tratar de equiparar las capacidades semánticas de humanos e IA, sino de si el resultado de sus funciones son equivalentes. De serlo, una máquina que reproduce actitudes doxásticas equivalentes a las que produciría un humano si éste estuviera en su lugar podría ser considerada como un igual epistémico, como trataremos de desarrollar en el siguiente epígrafe.

Alguien puede objetar que, en efecto, pueden reproducir comportamiento inteligente siguiendo un procedimiento de reglas bien definidas (Floridi 2024, pp. 39-43), pero eso no tiene nada que ver con una actitud doxástica genuina, sólo sobre seguir un algoritmo. Así es, una IA carece de estados mentales, al menos eso podemos afirmar ahora mismo. Puede darse el caso de que debido a un fenómeno emergente, una IA comience a generar creencias. Es

4 La IAD no opera con semántica (Searle 1983; 2000), pero sí entrega información semántica. Para que esto ocurra, sus datos se presentan en forma de enunciado declarativo, bien formado, significativo y veraz (Floridi 2013). En el caso de Maps, la ruta es la información semántica ofrecida por la IAD, y esta es trasladable desde un medio que presenta sus datos para ser vistos a otro en que se puede evaluar su verdad o falsedad, como es el caso de un enunciado (Floridi 2013 p. 186, Formigari 2004, pp. 91-92, Dretske 2000). Ese enunciado declarativo puede ser evaluado mientras que cumpla que los datos estén bien formados, sean significativos y veraces (Floridi 2013). "Bien formados" significa que se han ordenado según las reglas sintácticas del sistema utilizado. Por lo general, la sintaxis se relaciona con el lenguaje, pero su sentido puede ampliarse a cualquier sistema o código en el que los datos se agrupan siguiendo las reglas que los gobiernan. Datos "significativos" son aquellos que tienen significado en el sistema, código o lenguaje elegido. Que sean "veraces" implica que "proveen de contenidos verdaderos sobre el sistema modelado" (Floridi 2013, p.109). Esto último no significa que sean necesariamente verdaderos, solo que *pueden* actualizarse a verdaderos.

algo improbable, pero no imposible, pues otros fenómenos emergentes surgen causados por el hardware, como es la emisión de calor, ciertos zumbidos o incluso crujidos al ejecutar un programa (Searle 2000, p.26). Por el momento, no hay software, ni hardware que produzca inteligencia, ni remotamente parece que vaya a ser el caso en bastante tiempo (Floridi 2024), por lo que, consideramos, carece de sentido discutir esa posibilidad. Que la IAD produzca estados mentales genuinos o si solo los reproduce es de poca importancia para este artículo. Nos interesa en mayor medida lo que el comportamiento dice de la fiabilidad de sus procesos para producir actitudes doxásticas que sean estables y eficientes.

Aclaremos unas cuestiones relacionadas con Maps, ya que lo estamos tomando como ejemplo de IAD. Maps es una tecnología inteligente cuya IA está basada en aprendizaje automatizado [*machine learning*] y aprendizaje profundo [*deep learning*] (Lau 2020) que reproduce enunciados doxásticos sobre cuál es la mejor ruta para alcanzar un punto geográfico concreto. El método de aprendizaje es relevante porque determina de qué manera la IAD reproducirá el comportamiento inteligente. El aprendizaje profundo permite al dispositivo la toma de decisiones guiada por series de datos. Para que esto suceda, identifica y extrae patrones de esas series (Kelleher 2019) para después, desde esos patrones, ofrecer respuestas que se ajusten al marco de la solicitud de respuesta que demande una orden de entrada. De entre todas las respuestas probables, el dispositivo elige la que tenga un mayor índice de probabilidad de adecuarse al marco de la orden de entrada.

Maps es más complejo que un sistema de posicionamiento global (GPS). Todos los navegadores tienen GPS, pero no todos los GPS son navegadores inteligentes. El GPS relaciona el dispositivo con unas coordenadas y las proyecta sobre un mapa, pero no tendrá en consideración eventos novedosos que suceden en el trayecto. Maps recibe un comando de entrada para que el dispositivo trace una ruta óptima entre dos puntos. Esta IAD usa sus series históricas de registros de conductores humanos, desde donde extrae los patrones de comportamiento, siendo esa la parte *fija* de sus datos almacenados. A esto le añade una parte *dinámica* de series de datos, habitualmente aquellos que se están produciendo en el momento previo a la respuesta, pero que también sirven para actualizar el trazado durante el trayecto: geoposicionamiento de otros teléfonos que circulan por carretera, información del volumen de tráfico, accidentes, meteorología y condiciones del asfalto, entre muchas otras (Mehta, Kanani y Lande 2019, Lau 2020, Hassan 2022). Aunque la compañía responsable de Maps quiera situar su tasa de efectividad en un 97% (Lau 2020), siempre hay margen para el error.

Un conductor que conozca bien los alrededores, puede concluir que una ruta que parece peor en realidad es mejor al examinar evidencias que Maps es incapaz de calcular, como el comportamiento específico de los conductores que ocupan una vía, el tiempo que tardan los semáforos en cambiar de color, o que el vehículo va a pasar por una zona en donde la gente suele buscar aparcamiento, por lo que ese lugar tiende a generar retenciones. Maps tampoco es sensible al vehículo en el que se circula, por lo que desconoce cuestiones que probablemente el conductor tenga incorporada en su rutina, como que su coche se calienta en exceso en el centro de la ciudad, o que su motor le obliga a ir muy despacio en las cuestas.[5]

5 Estas cuestiones que afectan al portador del dispositivo son especialmente evidentes cuando Maps traza una ruta para un peatón cuyas especificidades son evitadas por el código, estableciendo el camino sin importar la condición física, edad o sus problemas cognitivos. Maps como IAD está diseñado fundamentalmente para la

3. La IAD como igual epistémico

Si la IAD cumple con las condiciones que exigimos a un ser humano para que lo consideremos como un igual epistémico, entonces deberíamos reconocer que la IAD también es un igual. A considera a B como igual epistémico, en caso de que B cumpla con las condiciones de

(1) igualdad probatoria, (2) igualdad cognitiva y (3) revelación completa.

Como condición previa, A debe sostener una actitud distinta al egoísmo epistémico y reconocer que hay otros seres con sus mismas capacidades epistémicas. La cuestión aquí es si B debe cumplir escrupulosamente con las tres condiciones o sólo en parte. Sostenemos que la igualdad absoluta es un ideal improbable que solo funciona desde una perspectiva teórica (Kelly 2005; King 2012; Cocchiaro y Frances 2021), pero eso no debe impedir que nos preguntemos que, en caso de que encontrar un caso de igualdad epistémica, cuáles son las condiciones que lo permiten. Defendemos que esas tres propiedades son simétricas, por lo que no puede haber un "término medio" (Zagzebski 2012, p.211): o bien aceptamos todas las condiciones o bien ninguna.

Consideramos que estas tres condiciones forman un mismo proceso de evaluación de la igualdad epistémica. Ninguna es suficiente por sí misma para demostrar que la IAD es un par epistémico. En (1) y (2) se evalúa si las propiedades y virtudes epistémicas del otro son simétricas, mientras que (3) tiene la función de presentar una justificación adecuada de cómo las propiedades y virtudes han elaborado *esa* creencia en lugar de otra cualquiera. En una situación de igualdad, (3) será simétrica, pues ambas partes llegarán a la misma creencia siguiendo el mismo camino de argumentos, razones o justificaciones. Hay que pensar que el proceso tiene que ocurrir de forma simultánea, es decir, no es un programa en el que primero se evalúa (1) y si es correcta se pasa a (2), y después a (3). Que las hayamos ordenado no se debe confundir con que sean pasos a seguir.

3.1. Igualdad probatoria.[6]

La igualdad probatoria exige que A y B conozcan las evidencias y los argumentos que influyen en la cuestión de P (Leslie 2013, Audi 2013). Si aceptamos que las evidencias

conducción de vehículos, cuando se trata de rutas a pie se comporta como un GPS que confunde aceras con calzadas. Sobre algunas cuestiones que se resuelven en ese párrafo, ver la nota 8.

6 Los argumentos que se exponen tanto en 3.1, 3.2 y 3.3 se orientan en la dirección de casos factuales. Con el ejemplo de Maps, partimos de la discutible idea de que podemos determinar factualmente cuál es la mejor ruta posible, cuando es un caso similar a un enunciado sobre el gusto ("Juego de Tronos es la mejor serie del mundo") o un enunciado moral ("Es mejor decir la verdad en una situación así"), esto es, son evaluativos. Agradecemos al revisor/a anónimo/a esta importante puntualización. Si analizamos las condiciones desde una perspectiva evaluativa, en lugar de la factual, los resultados nos llevarían a la misma conclusión que llegamos aquí, esto es, que la IAD no puede ser un igual epistémico, pero la argumentación debería tener en cuenta otros factores que aquí desestimamos. En cualquier caso, creemos que afrontar esta perspectiva evaluativa necesitaría un mayor espacio de elaboración, por lo que deberá ser abordada en otra circunstancia.

deben tener un carácter público (Zagzebski 2012), entonces tanto el humano como la IAD pueden acceder en igualdad de condiciones a las evidencias, establecer relaciones y formar una proposición sobre si P.

Pensemos que las series históricas que utiliza la IAD entrenada mediante aprendizaje automatizado son enormes en términos de cantidad de datos. Desde estas series de datos, que han sido generadas por seres humanos, identificará los patrones que servirán para dar una respuesta a la solicitud requerida (Kelleher 2019); por ejemplo, de la serie de datos sobre el comportamiento de todos los conductores que han pasado por las carreteras que van al aeropuerto durante los dos últimos meses se pueden extraer patrones que servirán para que la IAD reproduzca una actitud doxástica. El navegador cuenta tanto con estas series de datos como otras series que se están produciendo a tiempo real: tráfico, pavimento, obras, accidentes, etc. Puede dar la impresión de que en un caso como el del taxista Javier éste carece de ciertas evidencias comparativamente con la IAD, cosa que es incorrecta. Si todos los datos *deben* ser públicos, entonces siempre habrá alguna manera de acceder a éstos. Si Javier desconoce el volumen de tráfico e incidencias de los vehículos en la carretera A24, puede visitar la página de la DGT, llamar a otro taxista que esté en ruta, escuchar las noticias, etc. y acceder a las evidencias necesarias para formarse una opinión.

Lo que difiere considerablemente es la capacidad de la IAD en cuestiones como el almacenamiento o la velocidad de acceso y respuesta si se compara con las de un humano, ya que la máquina casi siempre le supera en rendimiento. Para determinar la igualdad epistémica esto es irrelevante. Excepto alguna que otra discusión sobre este tipo de capacidades de los iguales deben estar también a la par (Frances y Matheson 2019), pocas veces se tiene en cuenta. Solo es relevante si las capacidades cognitivas están claramente impedidas (Christensen y Lackey 2013).

Por lo tanto, si aceptamos lo anterior, podemos decir que la IAD y el humano mantienen una simetría de igualdad probatoria.

3.2. Igualdad cognitiva.

La segunda condición exige que A y B sean igualmente virtuosos en su evaluación de las pruebas y argumentos que influyen en P. Las virtudes epistémicas, por tanto, son el foco de la discusión (Audi 2013, Rowlands 2017), siendo algunas de estas virtudes la inteligencia, la capacidad para razonar y la ausencia de sesgos (Rowland 2017, Kelly 2005), entre otras. Si debemos buscar esas virtudes en procesos internos que sean idénticos a los de un humano, entonces no puede haber simetría. La IAD carece de rasgos mentales individuales (Baehr 2011) que la coloquen en situación de simetría con un humano; tampoco tiene un cerebro humano, y aunque las redes neuronales del aprendizaje profundo imiten la estructura de las conexiones nerviosas y neuronales, su funcionamiento es diferente. En la IAD no hay inteligencia, ni conoce el significado de las proposiciones que reproduce, pues es una máquina que únicamente opera con la sintaxis. Además, debe estar sesgada, porque los sesgos son condición necesaria para evitar problemas habituales en estos sistemas como el del sobreajuste y el del infraajuste (Kelleher 2019, p.20). Sabemos que los procesos de razonamiento que llevan a la extracción de datos son similares tanto en humanos como en la IAD, pero

sería demasiado polémico por nuestra parte tratar de defender que son fenómenos idénticos. Tampoco creemos que lo sean.

Existe otra estrategia. Se sigue desde cierto escepticismo ante la condición de virtud cognitiva, que se puede inferir de las aportaciones de King (2012) y de Cocchiaro y Frances (2021). Dado que es imposible tener razones suficientes para asegurar que aquel que tratamos de evaluar como igual cumple escrupulosamente con las capacidades exigibles, el acento se coloca en la expresión del comportamiento como reflejo de la actitud doxástica. Para la condición de igualdad cognitiva debería importar menos la correspondencia perfecta entre las funciones que generan los procesos evaluativos, racionales, sesgos, etc., e importar más el comportamiento como expresión de las capacidades epistémicas.

Nadie defiende que la IA tenga estados mentales, pues no produce actitudes doxásticas genuinas, pero tampoco hay una exigencia lógica de que los tenga. Como ya se señaló, aquí no defenderemos que la tecnología inteligente tenga estados mentales, ni la actual, ni ninguna que esté en un futuro a largo plazo, pese a las dudas que algunos filósofos tratan sostener al respecto (Chalmers 2023). Queda en su campo demostrar esta posición. Solo es necesario que el resultado de sus funciones resulten equivalentes a las que se le exigen a un humano para que sea calificada como igual epistémico.

Si tuviéramos que comparar los estados internos de un humano que producen una actitud doxástica y los procesos algorítmicos que reproducen una intención doxástica, no hay simetría posible. En cambio, si nos fijamos solo en el resultado, todo cambia. Por ejemplo, si un individuo ve unos platos limpios tiene complicado afirmar si fue un humano o un lavaplatos quien terminó la tarea, dado que el resultado del comportamiento del lavaplatos es equivalente al de un humano. Pero si alguien lo estuviera observando no los confundiría, aún siendo idéntica la entrada (los platos sucios) y el resultado del proceso (los platos limpios). Trasladando la analogía a una IAD, si no sometemos a escrutinio los estados internos de la máquina, se podría concluir que se da una simetría entre las capacidades epistémicas que producen resultados equivalentes.

¿Responden estas capacidades a alguna virtud epistémica? No necesariamente. En los últimos años se ha podido observar un cambio de rumbo desde la capital importancia de las virtudes epistémicas en la igualdad, como la perspicacia, la honradez o el rigor (Gutting 1982), hasta verse relegadas a un segundo plano, cobrando mayor importancia los antecedentes del igual epistémico (Bistagnino 2011, Cocchiaro y Frances 2021). Antecedentes como que haya demostrado sobrada competencia en P o de si está al tanto de las particularidades de P (Leslie 2013, Audi 2013). Así, resulta menos importante si las funciones internas operan bajo el concepto de virtud mientras que produzca actitudes doxásticas de calidad. Si consideramos una virtud epistémica como producir de manera eficiente y sistemática creencias verdaderas (Driver 2000, p.126), y si la IAD ofrece creencias verdaderas eficientemente y de manera estable, entonces habría que conceder que existe simetría.[7]

Al rastrear los antecedentes se responde a cuestiones del tipo "¿esta persona suele generar creencias acertadas sobre P?" Se puede auditar el historial de efectividad sobre P tanto de una persona como de una IAD. Si, por ejemplo, Maps traza habitualmente rutas que Javier

7 Aunque en el texto parece asumirse que la calidad de la información de una IAD se juzga por ser verdadera, en realidad habría que decir que ésta se juzga por su veracidad. Al respecto, ver la nota la nota al pie 6.

también considera como óptimas, este último se puede formar una opinión sobre la efectividad de la IAD, probablemente otorgándole un grado de confianza alto. Aunque Javier sepa que el comportamiento de Maps está producido por funciones automatizadas puede resultar irrelevante en tanto que lo que observa como respuesta se ajusta a sus propias creencias, y esto le lleva a concluir que tiene capacidades epistémicas equivalentes. Puede aplicar una máxima, como propone Roland (2017), por la que B es el igual epistémico de A acerca de P si B tiene la misma probabilidad que A de tener razón acerca de P.

En conclusión, la IAD cumple la condición de igualdad cognitiva siempre y cuando: a) no se tenga en cuenta las funciones internas, pues lo relevante es que hay un proceso de cálculo sobre las evidencias y un resultado en forma de actitud doxástica que es equivalente a la de un ser humano; y b) que no consideremos como fundamental que la IAD reproduce actitudes doxásticas en lugar de creencias genuinas.

3.3. Revelación completa.

En la revelación completa A y B comparten todas sus pruebas y argumentos sobre P. En la revelación completa el otro muestra sus cartas, de forma que se averigue si iba de farol. Se comparte, pero no es necesario discutir las pruebas, razones o argumentos sobre P, pues el objetivo de la revelación es determinar si la IAD puede ser considerada un par y no si puede haber un desacuerdo. Es necesario subrayar esto porque hay quien considera que la revelación completa sólo es necesaria para que surja un desacuerdo, pero no lo es para reconocer al otro como igual epistémico (Bistagnino 2011, p. 416), lo que implica que (1) y (2) son independientes de (3), siendo (3) irrelevante para la evaluación del par epistémico. Si un agente A considera que la IAD cumple (1) y (2), la situación de plena revelación no añade nada nuevo, pues la IAD ya es un igual. Estamos en sintonía con que la revelación completa marca el camino a seguir para determinar qué tipo de desacuerdo se está teniendo o si se produce un desacuerdo genuino, pero no coincidimos en que (3) sea una parte independiente que solo tenga esa función. De serlo, (1) y (2) serían suficientes para determinar si IAD es un igual epistémico, cosa con la que disentimos. Con (1) y (2) evaluamos las propiedades de nuestro posible igual, mientras que (3) tiene la función de presentar una justificación adecuada de cómo esas propiedades generan una creencia concreta. La simetría debe ampliarse más allá de las propiedades cognitivas. Debe mostrar cómo se ha formado una creencia concreta desde las evidencias hasta lo que se está afirmando, permitiéndonos también rastrear los antecedentes epistémicos de la otra persona.

Al ampliar la simetría con (3), encontramos tres elementos que ponen en cuestión que la IAD sea un igual epistémico: La ausencia de inteligibilidad; la IAD desconoce sobre lo que responde; la IAD no afirma nada.

*Inteligibilidad. El sistema de entrenamiento profundo es tan opaco que es prácticamente imposible rastrear cómo una IA toma sus decisiones. Es el problema de la inteligibilidad (Floridi y Cowls 2021, p.540), que también podríamos llamar de la explicabilidad o transparencia. Para que las explicaciones de una IAD sean tomadas en consideración debemos saber cómo funciona. No se trata de subrayar las diferencias entre las funciones de la máquina y el humano a la hora de hacer públicas sus razones, sino que debemos entender

cómo actuaría la IAD si estuviera en nuestro lugar. Si, por ejemplo, Maps tiene acceso a la información de que el coche se calienta en las curvas, y el navegador guía el vehículo con autonomía plena, ¿mantendrá ese valor como relevante al determinar la ruta como haría la persona propietaria del automóvil? El problema a día de hoy es que no podemos interpretar adecuadamente el proceso de toma de decisiones. Observar los resultados de la reproducción de comportamiento es insuficiente para justificarlo. Que a estos procesos se les considere una "caja negra" (Castelvecchi 2016) es sintomático de lo ininteligible de los comportamientos de estos sistemas. Incluso abrir la caja negra tampoco garantiza que lleguemos a comprender la toma de decisiones. La IA no representa la realidad mediante modelos matemáticos, sino que los crea desde sus series de datos (O'neil 2017), por lo que no solo deberíamos comprender qué camino tomó en su decisión, sino cuál es su modelo. Esto afecta a (1) porque la IA no va a mostrar cuáles son las evidencias que ha tomado y cómo las sitúa en una relevancia valorativa. El proceso es tan opaco en términos de inteligibilidad como lo es un código QR para un ser humano, sabemos que contiene información semántica, pero no la vamos podemos interpretar (Floridi 2013).

El problema de la inteligibilidad también se relaciona con la pregunta de "¿quién es el que responde?". Porque es el sistema el que lo hace, sí, pero la actitud doxástica reproducida no se corresponde con alguien concreto, sino que es la amalgama de comportamientos factorizados en un programa. No hay un alguien, sino un revuelto de actitudes. La evaluación de un igual debe darse sobre alguien concreto que sostiene una afirmación (King 2012; Cocchiaro y Frances 2021). Esto es diferente a que tenga que estar necesariamente presente en el momento de la afirmación. Se puede considerar un igual a alguien desde un testimonio, en el que se exponen las razones por las que ese alguien afirma algo sobre P. Siguiendo a Kelly (2013) y (Elga 2007), un catedrático que estudie la obra de David Hume le puede considerar como un igual, ya que entendemos que el proceso de demostración de las actitudes doxásticas de Hume quedaron registradas en sus escritos.[8] La igualdad no tiene que ser recíproca, por esa razón es irrelevante que la IAD nos reconozca como iguales, del mismo modo que Hume, dada su condición, tampoco lo hará.

La respuesta de la IAD de Maps, en el caso de Javier, es producto de un cálculo de probabilidad que se fundamenta en las decisiones de los conductores que pasaron por los caminos que llevan al aeropuerto. Dicho de otro modo, la IAD está produciendo respuestas vicariamente, en el sentido de que ésta sustituye sin autoridad a las opiniones registradas en sus bases de datos. Del mismo modo, siguiendo esta línea de razonamiento, si no se puede determinar quién debemos evaluar como igual, tampoco podemos considerarlo como una autoridad.

*La IAD no sabe lo que responde. En la segunda sección pusimos en valor que una de las objeciones sobre si la IAD puede ser un igual es que ésta carece de semántica o contenido mental alguno (Searle 1984, 2000). De este modo, una IA no sabe cuál es la mejor ruta, o qué es un cáncer, o quién escribió *Don Quijote*, aunque pueda trazar una ruta óptima,

8 Cappelen y Hawthorne (2009) diferencian entre "desacuerdo como actividad" y "desacuerdo como estado". Dos individuos pueden *estar* en desacuerdo sin *tener* un desacuerdo, como cuando no tienen contacto entre sí ni saben de las actitudes del otro. Por ejemplo, A es un catedrático de filosofía del siglo XXI y B es Hume, filósofo del siglo XVII. Si se diera la situación de considerar a la IAD como un igual y que esto nos lleve hasta un desacuerdo, estaríamos hablando aquí de *estar en desacuerdo*. Al respecto, Losada (2015).

señalar un cáncer en una tomografía, o atribuir a Cervantes la autoría del libro. Aunque de respuestas consistentemente eficientes, y de ahí su utilidad, acertar las respuestas es solo una parte de saber algo. Los loros también pueden responder de forma muy sofisticada a ciertas cuestiones, pero no afirmamos que sepan algo sobre lo que están graznando.[9] Es exigible ciertos procesos normativos que garanticen que esta sabe algo y no es simplemente azar afinado para asegurar la apariencia de éxito. Más allá de las reglas que se la hayan podido establecer durante su entrenamiento para que produzca resultados específicos, no existen normas que garanticen el valor epistemológico de los enunciados. En ocasiones ni siquiera es estable, ya que sobre la misma entrada de datos la IAD puede ofrecer resultados diferentes. Elige la opinión popular más aceptada, la que estadísticamente se ajusta mejor a los datos, en lugar de sostener una maquinaria robusta de generación de conocimiento. Si le exigimos que señale las fuentes de su respuesta, lo más cercano a una justificación será darnos una opción alternativa, referirse a su condición de máquina y que, por tanto, debemos tomar sus respuestas como "sugerencias", o apelar a la autoridad de alguna de sus fuente, dando muestra de que son incapaces de sostener sus propias afirmaciones y, por tanto, de la ausencia de compromiso hacia estas.

*La IAD no afirma. Sería atribuir demasiado a la IAD si decimos que afirma algo. Sus respuestas toman forma de un acto ilocutivo de afirmación, pero no puede ser considerada como tal. ¿Un acto locutivo? Puede ser, pero no debemos confundirlos con una afirmación genuina. Primero, porque carece de una consistencia entre lo que parece afirmar y su defensa posterior. Falta algo del acto de afirmar si el agente no se compromete con lo que dice sobre P. La respuesta de la IAD solo está garantizada desde las opiniones ajenas que se han estructurado sobre modelos matemáticos.

Segundo, la IAD incumple cualquier "norma de aserción", como aquellas que se refieren a la adecuación epistémica entre el hablante y el contenido de su afirmación (Pagin y Marsili, 2021). Incluso sucediendo que carecemos de consenso sobre si debe haber solo una norma o pueden ser varias, la IAD incumple todas aquellas que se consideran más robustas para explicar el acto de afirmar. La norma gobernada por conocimiento de Williamson dice que "hay que afirmar que *p* sólo si se conoce *p*" (2002, p.243). Si la IAD no conoce nada sobre *p*, no puede afirmar nada de *p*. De la misma manera, al no haber creencias genuinas en el enunciado de la IAD, ésta tampoco cumple con otras normas del tipo "hay que afirmar que *p* sólo si se cree que *p*", o que "hay que afirmar que *p* sólo si se está epistémicamente justificado en creer que *p*".

---

9 Una de las discusiones más relevantes sobre los Large Language Models (LLM) como ChatGPT, es si este es un "loro estocástico" (Bender, Gebru, McMillan-Major y Shmitchell, 2021). Un loro que acierta en la mayor parte de las ocasiones, sobre todo en cuestiones factuales, pero cuya arquitectura depende enormemente de las series de datos utilizadas en su entrenamiento. Otros no opinan igual y consideran que no se puede afirmar que estos modelos reproducen sus enunciados de forma caótica, pues las conversaciones que uno puede mantener en lenguaje natural alcanzan un nivel de fluidez que estadísticamente pueden considerarse como ordenadas (Arkoudas 2023). Sin embargo, esto solo oculta otros problemas subyacentes mucho más complicados de resolver, como la coherencia, esto es, mantenerse en el tema de la pregunta, producir sentido, adherirse a las máximas griceanas de conversación y demostrar un mínimo de entendimiento sobre cómo el mundo funciona (Arkoudas 2023, p.53). Esto nos lleva a que en una conversación prolongada con una IA ésta es incapaz de sostener sus propias afirmaciones.

En conclusión, si se revisa (1) y (2) desde (3) podemos concluir que aunque las primeras dos siguen siendo efectivas para determinar la igualdad con la IAD, su descripción es insuficiente. Podemos seguir defendiendo que sus procesos son equivalentes a los de un humano, pero detectamos que hay otros, como que no existe marco normativo alguno por el que garantizar la corrección de sus enunciados, que delatan la imposibilidad de que la IAD sea un igual epistémico.

4. ¿Al menos nos quedará la IAD como autoridad epistémica?

Un igual epistémico siempre es un tipo de autoridad epistémica, pero no al revés. Así que aunque no se pueda defender a la IAD como un igual, ¿al menos se puede conservar su estatus de autoridad? Si utilizamos la misma línea de razonamiento que se ha aplicado como objeción a que la IAD sea un igual epistémico, encontramos que el mismo razonamiento sirve para desechar que la IAD sea una autoridad.

Hasta ahora, hemos presentado a la IAD como medidor de lo verdadero. Un termostato es un buen medidor de la temperatura de la sala y probablemente sea la autoridad al respecto. El medidor de verdad es identificado por Zagzebski (2012) como autoridad epistémica débil o autoridad del experto, aquella que se circunscribe a una parcela concreta de la realidad. Más allá de los límites de su área de experto, deja de ser autoridad. Por último, los objetos inanimados pueden ser considerados autoridades débiles (Zagzebski 2012, p.119). Un dispositivo de GPS es una autoridad en situar a gente en unas coordenadas de geolocalización, pero solo si es eficiente en esa tarea. Las autoridades epistémicas débiles aportan creencias robustas a las que damos preferencia sobre las nuestras, ya que "el hecho de que la autoridad tenga una creencia P es una razón para que yo crea P, que sustituye a mis otras razones relevantes para creer P y no se añade simplemente a ellas" (p. 105, traducción propia). Por ejemplo, alguien cree que la lejía cura la COVID, pero si el sistema nacional de salud recomienda no hacerlo, y esta persona considera que el sistema nacional es una autoridad, abandonará su creencia y tomará partido por lo que los expertos sugieren (Jäger 2016). En este sentido, la autoridad débil puede proporcionar creencias de mayor calidad que las nuestras. Y eso que la persona del ejemplo no conoce a alguien concreto dentro del sistema de salud, o viceversa, nadie del sistema de salud conoce a esta persona.

La autoridad débil se diferencia de la auténtica autoridad en que la segunda requiere una relación interpersonal. Es equivalente a la autoridad religiosa o política: puede ordenar que se crea que P, sustituyendo así la creencia que tenga el agente. Un médico de familia es una autoridad epistémica por el crédito que alguien le otorga y va más allá de su actividad profesional. De la misma manera, una reseña en Amazon puede formar la creencia "no comprar un libro", pero si un amigo, del que siempre se valora su opinión, recomienda el libro, logrará que se sustituya la creencia sobre "no comprarlo" por "comprarlo" solo porque confiamos en este amigo. No está circunscrita a un espacio determinado o a un área del conocimiento, sino que es la relación entre sujetos la que establece la autoridad.

Entonces, dado que Maps no es un igual, ¿al menos podemos usar esa IAD para obtener creencias de mayor calidad? En cuanto a autoridad auténtica, tendríamos que descartar

que lo sea. Maps no pide que creas que su respuesta porque lo ordene. Tampoco establece una relación interpersonal en la que exista alguna responsabilidad entre las partes (Zagzebski 2012, p.119). La relación entre Javier y Maps es únicamente instrumental. Sería una irresponsabilidad por parte de Javier sustituir su creencia bajo la idea de que Maps es una autoridad. Pero, ¿y como autoridad débil? Maps ha demostrado ser efectivo en el trazado de rutas, por lo que Javier, que quiere estar en lo cierto, puede sustituir su creencia por la que ofrezca el dispositivo y depositar así su confianza en la fiabilidad del navegador.

Consideramos que hay una confusión en el razonamiento de Javier. Maps no es solo un GPS. Siguiendo a Zagzensbki, un GPS puede ser una autoridad en tanto experto en localizar a un sujeto en unas coordenadas y marcar rutas que tengan solo en consideración la distancia que separa el punto de salida y el de llegada. Maps geolocaliza, por lo que está situado en esa franja de experto, pero también reproduce una opinión sobre cuál es la mejor ruta, atendiendo a su serie de datos sobre el comportamiento de los conductores, así como los datos a tiempo real de flujo de tráfico, pavimento, climatología, etc. También modifica esa ruta sobre la marcha en función del comportamiento de otros conductores y recomienda rutas más largas pero "más eficientes en consumo de energía", por ejemplo.

Son actitudes doxásticas que surgen no porque sea el mejor juicio posible, sino que obedece a lo que fue el comportamiento de otros conductores. Son opiniones con poco fundamento, que son la media estadística de sus bases de datos. Una autoridad lo es porque es consistentemente fiable, pero también porque puede exponer razones (y defenderlas dado el caso) sobre P y tiene reconocida competencia sobre P ("sabe" sobre P). Es eficiente, pero también fiable. Podríamos suponer que Maps, como GPS, es eficiente, pero no es fiable. Por tanto, también es un error confiar en una IAD como autoridad epistémica. Es un instrumento doxástico que calcula cuál es la mejor opinión popular, pero que es incapaz de señalar cuáles son las razones que guían su respuesta.

5. Conclusiones

Hemos llegado a la conclusión de que una IAD no puede ser un igual epistémico. Aunque reproduce actitudes doxásticas y sus propiedades cognitivas sean equivalentes a las de un ser humano, algo que demuestra su comportamiento y resultados, en la revelación completa tenemos que considerar que estas propiedades son insuficientes. Para evaluar a alguien (o algo) como un igual epistémico debemos saber quién responde y así valorar su razonamiento, antecedentes y compromiso conceptual con lo que afirma sobre P. Una IAD no afirma nada, no sabe nada, no se compromete con sus enunciados y no se reconoce a quién corresponde la actitud doxástica que ésta reproduce. Estas mismas razones que nos llevan a negar que una IAD sea un par epistémico son igualmente aplicables para negar que sea una autoridad epistémica. Sería un error depositar nuestra confianza de manera incondicional en una IAD como autoridad epistémica débil, pues sólo algunas se fundamentan en comportamiento experto. Sería un gravísimo error considerarla como autoridad auténtica.

De entre las ventajas de conceder crédito a una autoridad epistémica es que, habitualmente, ésta nos proporciona creencias de mejor calidad. Análogamente, un igual epistémico, en caso de desacuerdo, nos podría empujar a que podamos re-evaluar nuestras creencias en

busca de otras de mejor calidad, justificación, robustez, etc. Dado que la IAD no es ni una cosa ni la otra, no podremos servirnos de esta ventaja.

Dicho esto, la calidad de la actitud doxástica reproducida por la IAD es relevante en cierta medida. Las actitudes doxásticas de la IAD no se forman de manera completamente azarosa. Depende del ajuste de los patrones de comportamiento de millones de datos que el aprendizaje automatizado aprovecha para construir las funciones de los modelos de las redes neuronales. Cuando la estadística y la programación son sólidas, así como el entrenamiento está afinado, podemos encontrar una IAD que funciona de forma considerablemente efectiva como generadora de opiniones. Es suficiente para reconocer a estas tecnologías como algo que podríamos llamar *igual doxástico*, dado que el resultado de su función es vicaria, esto es, está en lugar de una medida estadística de las actitudes doxásticas de humanos que fueron codificadas como series de datos. Aunque siga siendo un error confiar en la efectividad de la IAD como igual o como autoridad por las razones señaladas, podemos reconocer que la IA reproduce creencias estadísticamente más cercanas a lo correcto, sin que sea un efecto de "loro estocástico" o azar incoherente.

Referencias bibliográficas

Arkoudas, K. (2023) ChatGPT is no Stochastic Parrot. But it also Claims that 1 is Greater than 1. *Philos. Technol*. 36, 54. doi:10.1007/s13347-023-00619-6.

Audi, R. (2013). Dimensions of Intellectual Diversity and the Resolution of Disagreements. *The Epistemology of Disagreement: New Essays*, 205.

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. OUP Oxford.

Bender, E. M., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. En *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency,* 610-623.

Bistagnino, G. (2011). Epistemology of disagreement: Mapping the debate. *Gli annali di LPF-Laboratorio di Politica comparata e Filosofia Pubblica*, *6*, 159-187.

Cappelen, H. y Hawthorne, J. (2009) *Relativism and Monadic Truth.* Oxford. doi:10.1093/acprof:oso/9780199560554.001.0001.

Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, *538*(7623), 20.

Chalmers, D. J. (2023). Could a large language model be conscious?. arXiv preprint arXiv:2303.07103.

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

Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, *4*(5), 756-767.

Cocchiaro, M. Z. y Frances, B. (2021). Epistemically different epistemic peers. *Topoi*, *40*, 1063-1073. doi:10.1007/s11245-019-09678-x

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. MIT press.

Dretske, F. I. (2000). *Perception, knowledge and belief: selected essays*. Cambridge University Press.

Driver, J. (2000). Moral and epistemic virtue. *Knowledge, Belief, and Character: Readings in Virtue Epistemology, Lanham, MD: Rowman & Littlefield*, 123-34.

Elga, A. (2007). Reflection and disagreement. *Noûs*, *41*(3), 478-502.

Feldman, R. (2006). Epistemological puzzles about disagreement. *Epistemology futures*, 216, 236.

Floridi, L. (2013). *The philosophy of information*. OUP Oxford.

Floridi, L. (2023). *The Ethics of Artificial Intelligence: principles, challenges, and opportunities*. OUP Oxford.

Floridi, L. y Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.

Frances, B. (2014). *Disagreement*. John Wiley & Sons.

Frances, B. y Matheson, J., "Disagreement", *The Stanford Encyclopedia of Philosophy* (2019 Invierno), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/disagreement/>.

Gutting, G. (1982) *Religious Belief and Religious Skepticism*. Notre Dame: University of Notre Dame Press.

Hassan, H. (2022, 4 de enero). Google Maps: A simple explanation on how it detects traffic jams.https://medium.com/technology-hits/google-maps-a-simple-explanation-on-how-it-detects-traffic-jams-ce6940489c9c

Jäger, C. (2016). Epistemic Authority, preemptive reasons, and understanding. *Episteme*, 13(2), 167–185. doi:10.1017/epi.2015.38

Kelleher, J. D. (2019). *Deep learning*. MIT press.

Kelly, T. (2005). The epistemic significance of disagreement. *Oxford studies in epistemology*, *1,* 167-196.

Kelly, T. (2013). Disagreement and the Burdens of Judgment. *The epistemology of disagreement*: New essays, 31-53.

King, N. L. (2012). Disagreement: What's the problem? Or a good peer is hard to find. *Philosophy and Phenomenological Research*, 85(2), 249-272.

Lackey, J. A. (2013). Disagreement and belief dependence: Why numbers matter. In *The epistemology of disagreement: New essays* (pp. 243-268). Oxford University Press.

Lau, J. (2020, 3 de septiembre). Google Maps 101: How AI helps predict traffic and determine Routes. Google. blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/

Mehta, H., Kanani, P. y Lande, P. (2019). Google maps. *International Journal of Computer Applications*, *178*(8), 41-46.

O'neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pagin, P. y Marsili, N. (2021 Invierno), "Assertion", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/assertion/>.

Rowland, R. (2017). The epistemology of moral disagreement. *Philosophy Compass*, *12*(2), e12398. doi: 10.1111/phc3.12398.

Searle, J. R. (1982). The Chinese room revisited. *Behavioral and brain sciences*, 5(2), 345-348.

Searle, J. R. (2000) *El misterio de la conciencia*. Padios. Madrid.
Williamson, T. (2002). *Knowledge and its Limits*. Oxford University Press, USA. doi:10.1093/019925656X.001.0001
Zagzebski, L. T. (2012). *Epistemic authority: A theory of trust, authority, and autonomy in belief*. Oxford University Press.

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 137-152
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.612051

Plataformización, automatización y aceleración en los medios sociales

Platformization, automation and acceleration in social media

*RAÚL TABARÉS GUTIÉRREZ**

Resumen: La etapa de la "Web 2.0" despertó numerosas ilusiones sobre el potencial de los medios sociales para renovar y extender los espacios de deliberación, discusión y emancipación política de la ciudadanía. Sin embargo, la concentración empresarial que siguió al establecimiento de unas pocas plataformas como intermediarios culturales en el espacio digital ha propiciado diferentes reacciones en contra. Este ensayo presenta los principales problemas de los medios sociales para la discusión y deliberación. En particular, el texto aborda tres factores constituyentes de los mismos: la plataformización, la automatización y la aceleración. Se argumenta que estos factores presentan grandes dificultades para el desarrollo de espacios de deliberación en la red.
Palabras clave: Plataformización, inteligencia artificial, desinformación, censura, extremismo online, moderación de contenidos.

Abstract: "Web 2.0" period raised several expectations about the potential of social media to renew and extend spaces for deliberation, discussion and political emancipation of citizenship. However, business concentration that followed the establishment of a few platforms as cultural intermediaries in the digital space has led to backlash from different fields. This essay presents the main problematics associated to social media for discussion and deliberation. In particular, three constituent factors are identified: platformization, automation and acceleration. It is argued that these factors present great difficulties for the development of deliberative spaces in online environments.
Keywords: Platformization, artificial intelligence, misinformation, censorship, online extremism, content moderation.

Recibido: 12/04/2024. Aceptado: 26/06/2024.

* Dr. Raúl Tabarés es investigador senior en la Fundación TECNALIA RESEARCH & INNOVATION donde trabaja en la intersección entre digitalización, política y cultura. También es profesor asociado en el Máster de Retos Filosóficos de la Universitat Oberta de Catalunya. En la actualidad, su investigación está estrechamente relacionada con las culturas digitales y la innovación responsable.

1. Introducción

La aparición de diferentes medios sociales (del inglés "*social media*") tales como blogs, wikis y redes sociales, durante el periplo conocido como "Web 2.0"[1] despertó diferentes ilusiones y esperanzas sobre el potencial de este tipo de entornos online para renovar y extender los espacios de deliberación, discusión y emancipación política de la ciudadanía a través de las tecnologías digitales (Shirky, 2008; Surowiecki, 2005; Tapscott & Williams, 2006). Sin embargo, y pasado el optimismo inicial asociado a esta etapa, la concentración empresarial que favoreció el establecimiento de unas pocas plataformas como intermediarios culturales en el espacio digital ha propiciado numerosas reacciones en contra desde diferentes ámbitos (Gillespie, 2018; Poell et al., 2021; York, 2022).

Esta concentración empresarial se vio facilitada por la "Gran Recesión" y los grandes flujos de capital que se articularon en favor de los incipientes servicios basados en la Web, tales como motores de búsqueda, comercios electrónicos, etc., los cuales consolidaron un paradigma empresarial basado en negocios de intermediación apoyados en las posibilidades de Internet, comúnmente conocidos como "economía de plataformas" (Srnicek, 2017; van Dijck et al., 2018). Entre estos servicios se encuentran espacios de comunicación, expresión y discusión online (redes sociales, blogs o wikis) que monetizan la atención que depositan sus usuarios en diferentes contenidos, ostentando actualmente posiciones oligopolísticas o monopolísticas a través de empresas como Facebook, Twitter, Instagram o YouTube.

Estos medios sociales han estado involucrados en diversos escándalos asociados a la compañía Cambridge Analytica, cuyas actividades han influido en la debilitación de procesos democráticos, referéndums y elecciones, en Trinidad y Tobago, Reino Unido y Estados Unidos. Dichos episodios han puesto de relevancia el papel que los medios sociales ejercen como una amenaza real para la democracia (Bail, 2022; Gillespie, 2018; Poell et al., 2021; York, 2022). La percepción de los medios sociales ha ido evolucionando desde un optimismo inicial asociado a la etapa de la "Web 2.0" (Shirky, 2008; Tabarés, 2021; van Dijck, 2013), hacia una preocupación por la proliferación de la desinformación (Muirhead & Rosenblum, 2019), la creación de burbujas de opinión (Pariser, 2011), la radicalización y el extremismo online (Bail, 2022) o la censura (Roberts, 2019; York, 2022). Más recientemente, el desarrollo de nuevas formas de automatización en la creación de contenidos en el espacio online a través dela inteligencia artificial generativa (IAG) mediante "Large Language Models[2]" (LLMs) ha permitido el desarrollo de innovaciones como ChatGPT, Gemini o Co-Pilot. Estas innovaciones presentan nuevos retos y amenazan con reforzar las problemáticas asociadas a los medios sociales (Stahl & Eke, 2024; Franganillo, 2023).

A través de una revisión narrativa de la literatura, este ensayo aborda tres características de la idiosincrasia de los entornos deliberativos online: la plataformización, la automatización y la aceleración como factores constituyentes de los medios sociales, que presentan

1 Etapa que normalmente se suele contextualizar entre la crisis de las punto.com (finales de los años 90 y comienzos del siglo XXI) hasta el comienzo de la "Gran Recesión" (entre el 2008-2009).

2 Traducido en castellano como "Modelo de Lenguaje Regresivo" o "Modelo de Lenguaje Grande", los LLMs son sistemas de IA que predicen la siguiente palabra o carácter en un documento, y que basan su entrenamiento en una red neuronal con miles de millones de parámetros y grandes cantidades de datos sin etiquetar y de manera no supervisada.

grandes dificultades para el desarrollo de espacios de deliberación en la red. En el texto se aborda cada uno de estos factores por separado y se argumenta que el actual desarrollo de la IAG, focalizado en unas pocas plataformas y enmarcado en una lógica de "carrera armamentística", contribuye a fortalecer y extender las problemáticas asociadas a los medios sociales, inhabilitando en muchos casos los entornos online como espacio de reflexión y discusión.

El texto se articula de la siguiente manera: la siguiente sección identifica las tres problemáticas identificadas anteriormente. A continuación, el texto se centra en el análisis de los tres factores constituyentes del desarrollo de los medios sociales; plataformización, automatización y aceleración. Posteriormente se analiza cómo estos tres factores presentan formidables barreras para la discusión y reflexión en los entornos online, antes de explicar por qué la IAG contribuirá a empeorar la situación y cerrar el texto con unas breves conclusiones.

2. Problemáticas asociadas a los medios sociales

Si bien la aparición de los medios sociales como blogs, wikis y redes sociales fue aplaudida y ensalzada durante el periodo de "Web 2.0" como una oportunidad para renovar y extender los espacios de deliberación, discusión y emancipación política de la ciudadanía a través de las tecnologías digitales (Shirky, 2008; Surowiecki, 2005; Tapscott & Williams, 2006), lo cierto es que pasado este optimismo inicial, los medios sociales han sido criticados ampliamente por diversas problemáticas asociadas a su desarrollo y operativa.

Uno de los primeros problemas a los que se han asociado los medios sociales es la habilitación y el fomento de la desinformación (Bail, 2022; Gillespie, 2018), la cual está íntimamente relacionada con las teorías de la conspiración clásicas en las que se tratan de proveer de "explicaciones alternativas" a hechos controvertidos desde el punto de vista político. Esta desinformación más clásica se ha visto renovada por las nuevas oportunidades que ofrecen los medios sociales, las cuales no se basan en explicaciones alternativas, sino en el apoyo a diferentes teorías e ideas a través del apoyo de un gran número de usuarios y/o bots que contribuyen a su difusión y promoción a través de la creación y distribución de contenido relacionado (Muirhead & Rosenblum, 2019).

Además, el desarrollo de los medios sociales se ha asociado comúnmente al establecimiento de "filtros de burbuja" (Poell et al., 2021; Vaidhyanathan, 2018). La creciente personalización que se produce en los medios sociales, permite ofrecer resultados personalizados a diferentes usuarios de manera simultánea. Esta creciente personalización del contenido posee la contrapartida de que se evita la exposición del usuario a información que no está alineada con sus intereses y preferencias, reforzando sus puntos de vista y posiciones ideológicas, sin que éste entre en contacto con información contraria a sus ideas y valores (Pariser, 2011); una exposición necesaria en democracia, pues supone la aceptación y el respeto de ideas y posiciones ideológicas diversas. De igual manera, el riesgo a que los usuarios sean manipulados crece con los filtros de burbuja. Estos son utilizados por campañas de desinformación intencionadas que persiguen influir a la ciudadanía en diferentes procesos políticos. La normalización del extremismo online representa otra problemática de los medios sociales. (Bail, 2022). La radicalización de posturas se produce través del intercambio de opiniones y mensajes con usuarios que comparten posiciones, y al mismo tiempo, tienden a ver a sus

contrarios de un modo más alejado de sus posiciones de lo que realmente están. Este extremismo puede conllevar también a una "falsa polarización", es decir la sobrerrepresentación en el espacio online de minorías extremistas que son muy visibles debido a una actividad reseñable mediante la creación de contenido, comentarios, publicación de fotos, etc., y la movilización de dicho contenido en los medios sociales. Colectivos de extrema derecha y/o negacionistas del cambio climático, además de promotores de teorías conspiranoicas constituyen algunos ejemplos de esta falsa polarización, que también provoca que usuarios con posiciones más moderadas desistan o reduzcan sus interacciones en los medios sociales debido al riesgo de ser vejado en los entornos online (Bail, 2022).

Finalmente, los medios sociales han tenido que lidiar con la moderación de contenidos (Gillespie, 2018; Roberts, 2019). Se trata de una problemática difícil de solucionar debido a la escalabilidad y dimensión de estas plataformas que usualmente incorporan a millones de usuarios por todo el mundo. Para atajar este problema se han desarrollado técnicas de moderación de contenidos automatizadas apoyadas en IA. Estas técnicas monitorizan palabras clave y/o imágenes que involucran expresiones ofensivas y/o imágenes indecorosas con relación a los términos de referencia que articulan las políticas de plataformas como Facebook o YouTube.

Pese a ello, los medios sociales han contribuido al desarrollo de una industria auxiliar precaria e infrarremunerada en torno a la moderación comercial de contenidos y que normalmente se encuentra localizada en países del sur global como Filipinas, Sudáfrica o la India (Perrigo, 2023; Roberts, 2019). Países que culturalmente guardan una cercanía con la lengua y cultura anglosajonas, y que son las dominantes en este tipo de medios sociales. Esta cercanía es necesaria para poder evaluar correctamente las particularidades del lenguaje, tales como dobles sentidos, expresiones populares, alegorías, etc. (Gillespie, 2018). En palabras de Sarah T. Roberts, los profesionales de la moderación de contenidos comercial" son remunerados para revisar los contenidos subidos a los medios sociales en nombre de las compañías que facilitan las aportaciones de sus usuarios. Su trabajo es evaluar y decidir si el contenido online generado por los usuarios debe mantenerse o borrarse" (Roberts, 2019, 1).

Esta definición parece indicar que estos profesionales afrontan el dilema de mantener o borrar contenidos, pero la realidad es más complicada ya que dichos contenidos se pueden mantener acotando su visibilidad, impidiendo respuestas, etc. Recientemente, diferentes autores han señalado críticamente la censura que ejercen los medios sociales en Internet a través de esta moderación de contenidos (Gillespie, 2018; Poell et al., 2021; York, 2022). Gracias al gran poder que han ido acumulando con el paso del tiempo, plataformas como Facebook o Twitter, se han convertido en los guardianes y vigilantes de la información que millones de usuarios suben a sus plataformas, ostentando la potestad de censurar e invisibilizar ciertos temas que pueden ir en contra de sus "términos de referencia", ser contrarios a sus intereses comerciales o que puedan crear problemas de índole geopolítica.

En los últimos años diversos medios sociales han implementado mecanismos de geoposicionamiento en sus contenidos para restringir el acceso a contenidos promovidos por disidentes en países como Pakistán o Egipto (York, 2022). Otro ejemplo de esta censura es la que se ha producido en torno a los pechos de las mujeres en plataformas como Facebook e Instagram, donde ha habido varias campañas por parte de usuarias para la abolición de esas restricciones de contenido que afectaban a imágenes de mujeres que practicaban la lactancia

materna (Gillespie, 2018). Una censura que tiene implicaciones culturales, como la inhibición de determinados comportamientos que pueden moldear a la sociedad hacia actitudes más conservadoras y al mismo tiempo promover estereotipos idealizados de género.

3. Tres factores constituyentes de los medios sociales

3.1. Plataformización

Las bases de la plataformización se asientan durante el periodo conocido como "Web 2.0", las cuales no son sólo tecnológicas, sino también empresariales, legales y culturales para el desarrollo de un tipo de organización empresarial que se conocerá posteriormente como "plataforma" (Gillespie, 2010). A través del desarrollo de "aplicaciones web", se promueve una mayor facilidad de uso e involucración del usuario no especializado[3] en la creación, edición y distribución de contenidos en la web, promoviendo activamente la figura del prosumidor (Gutiérrez, 2015; Tapscott & Williams, 2006), es decir, un usuario que consume y genera contenidos. Simultáneamente, se empieza a experimentar activamente con modelos de negocio alrededor de este nuevo tipo de usuarios, que surgen en los medios sociales, asentándose posteriormente diferentes estrategias de monetización de estos espacios online a través de diferentes fórmulas y cuyo valor principal recae en la capacidad de recoger, analizar y reutilizar los datos y metadatos que producen estos usuarios en las plataformas (Gutiérrez 2015; van Dijck, 2013). Estos espacios suelen tener un acceso y uso gratuito y ello conlleva la necesidad de innovar en los modelos de negocio, al igual que en las tecnologías que los hacen posibles.

Esta plataformización es clave a la hora de ejercer diferentes roles de intermediación en una amplitud y diversidad de sectores tales como la publicidad, la educación, la movilidad, el consumo, el turismo o el entretenimiento, entre otros (Poell et al., 2021; Srnicek, 2017). Estos roles de intermediación adoptan a su vez posiciones monopolísticas u oligopolistas, no sólo desde el punto de vista tecnológico o económico, sino también desde una dimensión cultural.Las plataformas digitales han pasado a ser referentes sociales para diversas generaciones, especialmente en cuestiones que tienen que ver con la identidad online, la exposición pública, el intercambio de ideas, la discusión, el debate y el acceso a la información (Van Dijck et al, 2018; Gillespie, 2018; York, 2022). Los medios sociales han conformado un espacio virtual que algunos autores han denominado "infoesfera" (Floridi, 2014), o, incluso, "tercer entorno" (Echeverría, 1999; Echeverría & Almendros, 2023) , para resaltar la importancia de un contexto informacional superpuesto al entorno físico o material, legal y social, reconfigurando el mundo en el que vivimos de manera radical. La acumulación de poder por parte de los medios sociales ha centrado el interés de una considerable literatura académica que recoge sus implicaciones en el plano cultural, político, económico, tecnológico y social (Gillespie, 2010, 2018; Poell et al., 2021; Srnicek, 2017; van Dijck et al., 2018; York, 2022).

3 Aquel que no dispone de conocimientos de diseño y programación web.

La plataformización se posibilita a su vez por una serie de desarrollos e infraestructuras tecnológicas orientadas a facilitar la interacción y generación de datos (de forma voluntaria) y metadatos[4] (de forma involuntaria) por parte de los usuarios, fomentando valores como la conectividad, la modularidad o la interoperabilidad en el desarrollo tecnológico (Helmond, 2015). Diversas tecnologías Web como Ajax, Flash, HTML5, CSS3 o Javascript soportan la plataformización, pero sobre todo nuevas formas de conexión e integración digital como la API[5] orientadas a generar RIAs[6] que transforman los servicios Web tal y cómo se conocen, incorporando elementos multimedia, abriendo nuevas posibilidades de interacción con los usuarios, aumentando el tiempo y dedicación de los usuarios en estos entornos y recolectando una mayor cantidad de datos y metadatos de los usuarios (Poell et al., 2021; Tabarés, 2018).

Pero esta plataformización no se limita a factores técnicos, sino que engloba también factores sociales, organizativos, corporativos, políticos y mediáticos con los que las plataformas digitales buscan un posicionamiento en la esfera pública como intermediarios neutrales y que no estaban presentes en anteriores formas de organización empresarial (Gillespie, 2010). Las plataformas digitales combinan las características de un mercado horizontal (por ejemplo, los efectos de red presentes en Internet) con la verticalidad y jerarquía propia de las empresas privadas. Durante la época de la Web 2.0, este modelo de plataforma adquirió una serie de connotaciones positivas desde un punto de vista social, al albor de otros fenómenos e ideas que surgen en paralelo como la "inteligencia colectiva" (Surowiecki, 2005; Shirky, 2008), la "economía colaborativa" (Sundararajan, 2016) o el "consumo colaborativo" (Tapscott & Williams, 2006).

Por último, esta plataformización facilita el desarrollo de activos críticos para los medios sociales, por ejemplo, los diversos algoritmos propietarios y tecnologías de IA que se posibilitan a través de los datos y metadatos recolectados y el desarrollo de nuevas estructuras de computación dedicadas en la nube (Poell et al., 2021; van Dijck et al., 2018).

3.2. Automatización

El éxito de los medios sociales en la Web y el número creciente de usuarios en dichos entornos conlleva una mayor automatización en los entornos online. A través del desarrollo e implementación de diferentes elementos que favorecen la automatización en el tratamiento de la información, principalmente del uso de diferentes algoritmos, se abre las puertas a la gestión automatizada de esta información. Los algoritmos hacen posible la gestión automatizada de los millones de usuarios que suelen albergar los medios sociales, facilitando y extendiendo, además, la captura, recopilación, procesamiento, gestión y reutilización de los datos y metadatos que los usuarios generan en la interacción con

4 Es decir, datos que describen a su vez otros datos, tales como la geolocalización de una fotografía, la fecha en que se realizó, etc.

5 Del inglés, "*Application Programming Interface*", es un interfaz de software que permite a diferentes formatos de software poder comunicarse y facilitar la transferencia de datos entre ellos.

6 Del inglés, *"Rich Internet Applications",* es una aplicación web que se caracteriza por poseer las características de las aplicaciones de escritorio tradicionales, mejorando la experiencia del usuario de La Web.

estos entornos, ya sea de forma voluntaria o involuntaria (Srnicek, 2017; Tabarés, 2018; van Dijck et al., 2018).

Los algoritmos constituyen activos muy valiosos, ya que pueden aprender por sí mismos e identificar nuevas vías de innovación en forma de conocimiento, oportunidades y servicios que pueden permitir a los medios sociales mantener su posición en el mercado y mejorar su competitividad (O´Neill, 2017). Así, el desarrollo de nuevas funcionalidades en estos entornos suele estar directamente relacionado con el análisis de datos y metadatos, donde el papel de estos algoritmos es crucial. Por ejemplo, servicios como Google Trends[7] permiten conocer cuáles son las tendencias de búsqueda en un periodo concreto para un país o región determinado, así como explorar las razones de dichos comportamientos en la población, para proveer soluciones que puedan dar respuesta a dichas demandas en un futuro cercano.

Al mismo tiempo, los algoritmos son importantes instrumentos de poder para los medios sociales. A través de las clasificaciones y categorías que se posibilitan en estos entornos, contribuyen a reorganizar el orden social (Noble, 2018). Esta recategorización es visible en plataformas como Twitter, donde a través de los "*hashtags*" o etiquetas, el usuario puede conocer cuáles son los temas que más atención recaban por parte de los usuarios en un determinado momento. Atención que estas plataformas son capaces de rentabilizar a través de publicidad contextualizada y segmentada en función del tipo de usuarios que acceden a estos temas de conversación, y que constituye la principal fuente de ingresos para medios sociales como Twitter o Facebook. Una publicidad cada vez más digitalizada y donde los grandes presupuestos de diferentes marcas de consumo explotan las posibilidades del entorno online, no sólo a través de estas compañías, sino de figuras emergentes en estos entornos como los "influencers".

Por otro lado, los algoritmos también juegan un papel crucial en diferentes tipos de plataformas orientadas a la discusión, debate y deliberación online. Con el desarrollo y popularización de diferentes tecnologías de "*Big Data*" e IA, la mayoría de las plataformas han ido implementando progresivamente técnicas de moderación automatizadas (Gillespie, 2018; Poell et al., 2021). Técnicas de IA que se basan en "Machine Learning" (aprendizaje supervisado por humanos) o "Deep Learning" (sin supervisión) y que son críticas para el desarrollo de herramientas automatizadas que puedan identificar lenguaje ofensivo, contenido pornográfico, noticias falsas, propaganda o materiales susceptibles de atentar contra la dignidad humana, tales como violaciones, actos terroristas, mutilaciones, etc.

El 19 de agosto de 2014 representa una fecha señalada en relación con estas herramientas automatizadas. En esa fecha la organización terrorista ISIS divulgó, en YouTube, un vídeo sobre la decapitación del periodista James Foley. Las imágenes dieron la vuelta al mundo en pocas horas, revelando la debilidad de los medios sociales para evitarlo y enfatizando la necesidad de desarrollar herramientas mucho más avanzadas para controlar la difusión de contenidos dañinos (Gillespie, 2018; Poell et al., 2021). El creciente escrutinio público al que se sometieron los medios sociales, también con motivo de las elecciones estadounidenses a la presidencia en 2016, y de otros escándalos que han salpicado posteriormente a varios medios sociales (por ejemplo, en los atentados de Cristchurch en Nueva Zelanda) han obligado a adoptar medidas conjuntas en el sector. Por un lado, se han incrementado

7 https://trends.google.es/trends/

los esfuerzos para desarrollar una industria auxiliar de moderación de contenidos a costa del trabajo humano infrarremunerado y precario en países del sur global. Por otro, las plataformas han implementado diferentes técnicas de moderación de contenidos automatizadas (Gorwa et al., 2020).

Las técnicas automatizadas explotan las posibilidades de las tecnologías de IA existentes para localizar palabras que vayan en contra de "términos de referencia" y políticas asociadas a determinadas plataformas, a través del procesamiento natural del lenguaje, o contrastando imágenes con bases de datos existentes, para intentar frenar la entrada de contenidos. Sin embargo, la opacidad legal (por derechos de propiedad intelectual de los algoritmos empleados), técnica (la ausencia de auditorías algorítmicas a estos sistemas) y de gobernanza (por cómo se formulan las políticas y términos de referencia que luego se ejecutan a través de los algoritmos) imposibilita una mayor transparencia en cómo se llevan a cabo estas prácticas de moderación automática de contenidos.[8] En los últimos años, además, ha habido presiones políticas de regímenes autoritarios que han condicionado la operativa de los medios sociales en ciertos países a diferentes políticas, añadiendo una mayor opacidad y complejidad a su funcionamiento (York 2022).

3.3. Aceleración

Junto a la plataformización y automatizaciones causadas por las estructuras organizativas y tecnológicas que se han descrito anteriormente, existe una tercera capa constituyente de los medios sociales: la aceleración provocada por la ruptura de la relación espacio-tiempo . Esta ruptura ha sido objeto de análisis por autores como Manuel Castells (1997). En su famosa trilogía sobre la era de la información, Castells explica cómo dichas tecnologías facilitan el acceso a la información en cualquier lugar y a cualquier hora, desdibujando los límites temporales y espaciales. Así mismo, también enfatiza la "constante conectividad" para referirse a la centralidad del entorno informacional digital (Echeverría, 1999; Echeverría & Almendros, 2023; Floridi, 2014).

Aquí empleamos el término "aceleración" para indicar que esta brecha espaciotemporal representa también un factor constituyente de los medios sociales La aceleración está muy presente en las dinámicas de interacción que se facilitan en estos espacios online, donde se incita al usuario a consumir diversos contenidos a través de los famosos "muros" (del inglés "*timeline*") y de ruletas de contenidos (el famoso "*scrolling*"), que imitan el funcionamiento de las máquinas tragaperras, con el objetivo de capturar la atención del usuario el máximo tiempo posible. Además, los medios sociales incitan comúnmente a los usuarios a compartir continuamente sus pensamientos e ideas (el famoso ¿qué está ocurriendo?), así como a reaccionar a los contenidos de otros usuarios (me gustas , favoritos o "*retuits*") y generar conversaciones alrededor de los contenidos de otros usuarios con inmediatez para que se puedan viralizar dichos contenidos (prueba de ello son los famosos concursos online donde se incita al usuario a participar a través de un comentario y otras acciones adicionales).

8 Ver por ejemplo https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms

El pensador alemán Hartmut Rosa (2016, 2019) ha prestado especial atención a la aceleración que se produce en el espacio online. Rosa ha dedicado considerables esfuerzos a desarrollar una teoría crítica de la temporalidad en la modernidad tardía, y a explicar cómo la aceleración social presenta grandes problemas para la realización de una buena vida, además de ser una considerable fuente de alienación social. Este autor documenta la velocidad e intensidad con la que se desarrollan los procesos sociales en la actualidad; dinámicas que provocan una crisis de la experiencia debido a que los individuos no tienen tiempo material para vivir plenamente las experiencias que abordan por el régimen temporal en el que se circunscriben. Esta aceleración causa una sensación de alienación y desapego, debido al consumo compulsivo de experiencias. Rosa identifica tres fuentes de aceleración: la tecnológica (referida al desarrollo de tecnologías que permiten cubrir mayores distancias y comunicarse más rápidamente) la del cambio social (principalmente en torno a las relaciones en el trabajo y la familia y su carácter cada vez menos estable) y la del ritmo de vida (por agregación de las otras dos y referida principalmente a la presión que sienten los individuos por hacer más cosas en menos tiempo en diferentes ámbitos).

Los medios sociales constituyen una gran fuente de aceleración tecnológica, pues introducen nuevas tecnologías e innovaciones en sus respectivos entornos para captar la atención de sus usuarios a través de estas nuevas funcionalidades y favorecer la generación, consumo y distribución de contenidos(Gutiérrez, 2015; Tabarés, 2021). Por este motivo, muchos de los medios sociales hoy en día disponen de grandes inversiones y programas de investigación en tecnologías emergentes como la IA o el metaverso[9].

Al mismo tiempo, también promueven la aceleración del cambio social, animando a los usuarios a compartir su privacidad, tanto en el ámbito familiar como el laboral. Así, por ejemplo, han surgido prácticas de "*sharenting*" en las que los progenitores crean narrativas de vida y desarrollo infantil mediante fotos, vídeos y/o textos. Esta exposición pública no está exenta de controversias y problemáticas y actualmente es también motivo de investigación por las implicaciones sociales y éticas respecto a menores. De igual modo, la exposición mediática de los logros individuales y/o colectivos en el entorno laboral también ha sido un tema recurrente en el desarrollo de los medios sociales, con el lanzamiento de aplicaciones sociales específicas para el entorno laboral, tales como Linkedin, Yammer o Slack. Exponer públicamente los diferentes logros laborales, así como mostrar el grado de ocupación que se atesora se ha convertido hoy en día en un símbolo de estatus (Wajcman, 2020). Sin embargo, también hay problemáticas asociadas a esta aspiración a estar constantemente ocupados, con la aparición del denominado síndrome de "*burnout*" (traducido como "síndrome del trabajador quemado") y que se erige como uno de los males más ilustrativos de la obsesión actual de la sociedad por la productividad y la eficiencia, en la que los medios sociales también juegan un papel importante (Han, 2012).

9 Ver por ejemplo https://about.meta.com/metaverse/

Por último, los medios sociales contribuyen innegablemente a la aceleración del ritmo de vida, no sólo por la exposición mediática a la que dotan a las experiencias de vida de sus usuarios, sino también porque contribuyen a un acelerado consumo de estas experiencias, íntimamente asociado al modelo de negocio en el que se basan y que no es otro que la publicidad hiper-segmentada (Gutiérrez, 2015; Poell et al., 2021; Tabarés, 2021; van Dijck et al., 2018). De esta manera, y a través de diferentes tecnologías basadas en los datos que se recogen y analizan en los medios sociales, se realizan asociaciones entre los contenidos que los usuarios de estos medios sociales publican, comentan, y/o comparten, con los servicios y productos que anuncian diferentes empresas y marcas comerciales. Por ello, los medios sociales se han convertido en un entorno que favorece la aceleración del consumo de productos y servicios que gozan de una gran exposición mediática en el entorno online.

4. ¿Discusión y deliberación en medios sociales?

Como se ha señalado en el texto, la plataformización, automatización y aceleración son tres factores constituyentes claves para el desarrollo, escalabilidad y consolidación actual de los medios sociales, posibilitando posiciones de oligopolio o monopolísticas decisivas para la congregación de miles de millones de usuarios en sus respectivos entornos informacionales. Sin embargo, estos tres factores constituyentes presentan significativas problemáticas para promover la deliberación y la discusión en el entorno online. A continuación, se exponen cuáles son las problemáticas que acarrean estos tres rasgos para la discusión y deliberación en los entornos online (tabla resumen 1).

En primer lugar, la plataformización es un factor limitante de las posibilidades para la discusión y deliberación en los entornos online, ya que los usuarios que deciden crear contenido, comentar contenido de otros, crear un grupo de debate o participar de una conversación, deben de hacerlo en torno a los términos de referencia que plantea el medio social en cuestión. El papel de estos términos de referencia siempre es controvertido, ya que son lo suficientemente abiertos para no restringir casi ningún tipo de ideal, pero al mismo tiempo su aplicación y puesta en práctica, a través de la moderación de contenidos, puede ser tremendamente restrictivo (Gillespie, 2010, 2018; Poell et al., 2021; York, 2022). Ciertos episodios como los mencionados anteriormente en el texto en torno a las imágenes de madres que tratan de fomentar la lactancia materna compartiendo fotografías mundanas y que han sido objeto de censura por parte de estos medios sociales, ilustran el poder que atesoran los medios sociales a la hora de definir los términos del debate y los temas a discutir.

Además, estos términos de referencia no son estables y sufren diferentes tipos de actualizaciones, revisiones y/o modificaciones con mucha frecuencia y de manera aleatoria, que responden a nuevas funcionalidades técnicas, requerimientos legales y/o diferentes controversias, pero que en ningún caso son explicados de manera apropiada a sus usuarios. Es decir, no ofrecen ningún tipo de marco normativo estable para los usuarios de estos servicios. Por otro lado, también es importante resaltar que las conversaciones que suceden en un medio social en particular, no se pueden extender más allá de los límites

de ese medio social en cuestión. La conversación global y fragmentada que se produce en Facebook no puede extenderse a Twitter, sino que la misma persona debe abrir diferentes perfiles en los diferentes medios sociales, mostrando un estilo de conversación diferente y adaptándose a las particularidades de los diferentes entornos, con el objetivo de maximizar la viralidad de dichos contenidos. Por ello, algunos autores mencionan como esta fragmentación del usuario (y la persona) en la Web da lugar a una fragmentación social (Vaidhyanathan, 2018) que ofrece diferentes versiones de uno mismo (Bail, 2022) y que presenta problemas ontológicos de cara a la representación de la persona en el medio online (Echeverría & Almendros, 2023).

En segundo lugar, la creciente automatización a la que asistimos en los medios sociales tiene varias implicaciones para una cultura de discusión y deliberación en los entornos online. Como ya hemos visto, el uso de herramientas de moderación de contenidos automatizadas es controvertido por el reduccionismo que se ejerce al limitar las problemáticas socioculturales que afloran en los medios sociales a la prohibición de ciertos términos y/o imágenes (Gillespie, 2018; Roberts, 2019). Este reduccionismo, unido a la falta de información sobre el empleo de estas técnicas y la opacidad que existe alrededor de ellas, dificulta la evaluación de su eficacia e imposibilita análisis externos que puedan delimitar hasta qué punto pueden ser sustitutivas de la moderación de contenidos humana. Así mismo, no hay instrumentos o instituciones externas que puedan auditar dichas prácticas (Gorwa et al, 2020).

El auge de la IA y la IAG ha visto como esta moderación de contenidos humana también se ha desplazado al etiquetado y filtrado de datos con las que se entrenan los LLMs (Perrigo, 2023). El objetivo de estas actuaciones consiste en prevenir que modelos de IAG como ChatGPT o Gemini reproduzcan lenguaje ofensivo, discriminatorio o los sesgos propios que podemos encontrar en multitud de páginas web donde la libertad de expresión es llevada a sus límites como Reddit, y donde muchos de los estereotipos, prejuicios y sesgos existentes en las sociedades occidentales son explicitados. Estos sitios web son también los que comúnmente se utilizan en el entrenamiento de estos sistemas, ya que por otro lado son ricos en estructuras lingüísticas comunes, vocabulario informal y gozan de una gran riqueza léxico-gramatical. Aspectos clave a la hora de que una IA pueda reproducir de una forma probabilística y fiable los aspectos comunicativos clave en una conversación con un usuario.

Por último, y quizás de manera más preocupante que respecto a los dos anteriores factores constituyentes, la aceleración que se promueve en este tipo de entornos online supone una gran barrera para el intercambio de opiniones de manera reposada y reflexiva. La lógica comercial que reside de manera subyacente a las infraestructuras tecnológicas posibilitadas por los medios sociales no recompensa los intercambios con dilación, sino con premura (Rosa, 2016). Así, el contenido que se viraliza y que se hace popular en un determinado medio social es aquel que atrae la atención de un mayor número de personas en un espacio de tiempo más reducido. Es decir, se premia la reacción e interacción mundana con el contenido, pero no la reflexividad, la crítica o la meditación. Funciones, todas ellas, que requieren de tiempo para pensar y actuar. Algo que los medios sociales no facilitan ni habilitan, con su continuo dinamismo en los muros de los usuarios, favoreciendo contenido recientemente creado por diferentes usuarios, que capturen la atención del mismo y favorezcan sus reacciones en forma de "me gustas" o "retuits" (Vaidhyanathan, 2018).

Factores constituyentes de los medios sociales	Problemáticas para la discusión y deliberación online
Plataformización	• Limitaciones del debate en torno a los términos de referencia y políticas específicas. • Inestabilidad en los marcos normativos del debate y censura puntual. • Compartimentación y fragmentación de la conversación y el debate.
Automatización	• Reduccionismo y limitaciones en la moderación de contenidos. • Opacidad y falta de transparencia en los instrumentos de moderación de contenidos. • Propagación de sesgos y estereotipos.
Aceleración	• Consumo acelerado y compulsivo de contenidos. • Carencia de espacios y tiempos para la reflexión. • Viralización y sobreexposición de contenidos emocionales.

Tabla 1. Problemáticas asociadas a los factores constituyentes de los medios sociales. Elaboración propia a partir de la revisión bibliográfica.

5. La IA, una intensificación de la plataformización, automatización y aceleración de los medios sociales

La popularización de los medios sociales y su creciente centralidad en la sociedad ha sido clave en el desarrollo de innovaciones de IA como Siri. Este tipo de innovaciones han sido posibilitadas por técnicas de IA basadas en "*deep learning*" que posibilitan que los sistemas de IA aprendan de una dieta de información de manera incontrolada y de múltiples tipos de datos (Tabarés, 2020). Este tipo de interfaces conversacionales también han sido objeto de una gran revolución a través de los LLMs y las arquitecturas de computación "*Transformer*" en las que se basan (Stahl & Eke, 2024). Modelos de IA que son entrenados en base a un ingente número de corpus lingüísticos procedentes de Internet y la Web, tales como la Wikipedia y que han sido claves para el desarrollo de innovaciones como ChatGPT (Open AI), Gemini (Google) o Copilot (Microsoft).

Estos chatbots interactúan con el usuario a través de "*prompts*" o mensajes de texto en los que se realizan consultas por parte del usuario, a partir de los cuales el sistema es capaz de generar textos, imágenes o incluso vídeos. Estos sistemas asignan una probabilidad a un determinado "*token*[10]" que puede ser un término o un pixel determinado. Su funcionamiento se basa en predecir el próximo término que seguirá a otro de manera probabilística, y de esta manera son capaces de generar textos largos, traducirlos a otros idiomas, resumirlos, etc. Sin embargo, como todo sistema basado en la probabilidad, no están libres de errores. En particular, estos sistemas cuentan con mensajes en su interfaz de usuario en los que se resalta que pueden proveer de información errónea al usuario, además de que heredan en

10 Un token es una referencia o un identificador que sustituye a un término de texto y que permite ser referenciado en un sistema de tokenización

muchas ocasiones los sesgos existentes en los corpus lingüísticos existentes en Internet y la Web, y en ciertos casos donde tienen incertidumbres probabilísticas, directamente se la inventan (las denominadas "alucinaciones").

Es de prever que la irrupción de la IAG en el entorno online contribuya a agravar las problemáticas asociadas a los medios sociales que hemos descrito anteriormente, tales como la desinformación, la creación de burbujas de opinión, la radicalización y el extremismo online o la censura (Franganillo, 2023). Estas nuevas innovaciones permiten el desarrollo de "*deepfakes*[11]" (vídeos falsos de personas que aparentemente son reales), facilitan el desarrollo de campañas de desinformación a gran escala por su capacidad de generación de textos e imágenes, pueden favorecer la manipulación y suplantación de identidad por el uso de lenguaje persuasivo, y en algunos casos convincente, así como la implementación de autocensura en estos sistemas en relación a diferentes términos controvertidos que directamente se eliminan para favorecer su uso por un número mayor de usuarios.

Sea como fuere, los tres factores constituyentes de los medios sociales que hemos analizado en este texto, ya se están viendo reforzados, y es muy probable que vayan a ser extendidos en el futuro. La reciente carrera armamentística en la que parece que se ha convertido el desarrollo de la IA, y de la IAG en particular, refuerza la lógica de la plataformización, ya que solamente unas pocas compañías en el mundo pueden permitirse desarrollar estos sistemas. Solo aquellas que disponen de la ingente cantidad de datos necesaria para su entrenamiento, de los costosos recursos computacionales que se requieren para su desarrollo y del personal especializado que desarrolla los algoritmos que procesan toda la información con la que se entrenan (Stahl & Eke, 2024). Al mismo tiempo, la automatización en los medios online también se verá reforzada con el uso de la IAG en ellos, ya que ChatGPT y otras innovaciones ofrecen diferentes potencialidades a la hora de automatizar ciertas tareas y su uso en las redacciones periodísticas y en el marketing ya es un hecho. Por último, la IAG permitirá una velocidad mayor a la hora de generar y difundir contenidos, ya sean de texto o multimedia, contribuyendo a incrementar la aceleración en los medios sociales (Franganillo, 2023).

Conclusiones

Este artículo ofrece un repasado por las principales dificultades que plantean los medios sociales para la discusión y deliberación online, desde la proliferación de la desinformación (Muirhead & Rosenblum, 2019), la creación de burbujas de opinión (Pariser, 2011) la radicalización y el extremismo online (Bail, 2022) o la censura (Gillespie, 2018; Roberts, 2019; York, 2022). A partir de estas problemáticas, el trabajo identifica factores constituyentes de los medios sociales –la plataformización, la automatización y la aceleración–, y los pone en relación con el cuestionamiento de que los medios sociales favorezcan una cultura de la discusión y deliberación. La fragmentación, la falta de transparencia, el reduccionismo, la inestabilidad y la falta de reflexibilidad representan los impedimentos fundamentales. El artículo argumenta que el desarrollo de la IA, y en particular de la IAG, refuerza la

11 El término "*deepfake*" proviene del inglés y está formado por la unión de "*deep learning*" (una técnica de IA de aprendizaje automático) y "*fake"* (falso).

plataformización,automatización y aceleración de los medios sociales deteriorando su potencial como espacios de reflexión y discusión online.

Como solución, se ha resaltado el papel de la regulación para promover prácticas más transparentes, auditables y responsables por parte de la industria de los medios sociales. A pesar de que la entrada en vigor de la "Digital Services Act"[12] supuso un primer paso, es necesario avanzar en esta línea para intervenir en un sector que posee una gran influencia cultural en la sociedad y dispone de grandes implicaciones para el futuro de nuestras sociedades democráticas.

Referencias

Bail, C. (2022). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.

Castells, M. (1997). *La Era de la Información. Vol I: La Sociedad Red*. Alianza.

Echeverría, J. (1999). *Los señores del aire: Telépolis y el tercer entorno*. Destino.

Echeverría, J., & Almendros, L. S. (2023). *Tecnopersonas: Cómo nos transforman las tecnologías*. Ediciones Trea.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. methaodos.revista de ciencias sociales, 11(2), m231102a10. http://dx.doi.org/10.17502/mrcs.v11i2.710

Gillespie, T. (2010). The Politics of Platforms. *New Media & Society*, *12*(3), 347–364. https://doi.org/10.1002/9781118321607.ch28

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data and Society*, *7*(1). https://doi.org/10.1177/2053951719897945

Gutiérrez, R. T. (2015). *La Belleza del Código: Influencia de la Web 2.0, los medios sociales y los contenidos multimedia en el desarrollo de HTML5* [Universidad de Salamanca]. http://dsp.tecnalia.com/handle/11556/190

Han, B.-C. (2012). *La sociedad del cansancio*. Herder.

Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*, *1*(2), 205630511560308. https://doi.org/10.1177/2056305115603080

Muirhead, R., & Rosenblum, N. L. (2019). *A lot of people are saying: The new conspiracism and the assault on democracy*. Princeton University Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

O´Neill, C. (2017). *Weapons of Math Destruction. How Big Data increases inequality and threatens democracy*. Penguin Books.

12 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Books.
Perrigo, B. (2023, January 18). *OpenAI Used Kenyan Workers on Less Than $2 Per Hour*. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/
Poell, T., Nieborg, D., & Duffy, B. E. (2021). *Platforms and cultural production.* Polity Press.
Roberts, S. T. (2019). *Behind the screen*. Yale University Press.
Rosa, H. (2016). *Alienación y aceleración. Hacia una crítica de la temporalidad en la modernidad tardía.* Katz Editores.
Rosa, H. (2019). *Remedio a la aceleración. Ensayos sobre la resonancia*. Ned ediciones.
Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. Penguin.
Srnicek, N. (2017). *Platform Capitalism*. Polity Press.
Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*. https://doi.org/10.1016/j.ijinfomgt.2023.102700
Sundararajan, A. (2016). *The Sharing Economy*. The MIT Press.
Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.
Tabarés, R. (2018). Understanding the role of digital commons in the web; The making of HTML5. *Telematics and Informatics*, *35*(5), 1438–1449. https://doi.org/10.1016/j.tele.2018.03.013
Tabarés, R. (2020). Conversando con cajas negras; sobre la aparición de los interfaces conversacionales. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, *17*(2), 179–186. https://dx.doi.org/10.5209/TEKN.69303
Tabarés, R. (2021). HTML5 and the evolution of HTML; tracing the origins of digital platforms. *Technology in Society*, *65*. https://doi.org/10.1016/j.techsoc.2021.101529
Tapscott, D., & Williams, A. D. (2006). *Wikinomics: la nueva economía de las multitudes inteligentes*. Paidos Ibérica.
Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy.* Oxford University Press.
van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press.
van Dijck, J., Poell, T., & Waal, M. de. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
Wajcman, J. (2020). *Pressed for time: The acceleration of life in digital capitalism.* University of Chicago Press.
York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism.* Verso.

SIMPOSIO SOBRE *WHO SHOULD WE BE ONLINE* (OUP, 2023) DE KAREN FROST-ARNOLD

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 155-156
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.620291

Précis of *Who Should We Be Online? A Social Epistemology for the Internet*

KAREN FROST-ARNOLD[*]

Who Should We Be Online? provides a socially situated epistemology for the internet. There are many important epistemological questions about the internet, and in recent years concerns have grown about the internet's effect on what we believe. Epistemologists have rapidly become interested in the problems of fake news, disinformation, conspiracy theories, and the role of large social media companies in shaping our media and public spaces for debate. This book builds on this literature, but it also argues that something important has been largely missing from extant philosophical analysis of the internet. Social epistemology needs to pay attention to the role of power, oppression, and inequality in shaping what we know and what we don't know online. Racism, misogyny, homophobia, transphobia, ableism, colonialism, and other forms of oppression both influence individual users' online behavior and also structure the platforms, policies, and design features of social media spaces. Prejudice often shapes who we trust and distrust online, and structural oppression affects whether online platforms can be reasonably trusted by marginalized people. Whom we trust has immediate consequences for what we do and do not know. Unfortunately, much social epistemology of the internet abstracts away from the social context of knowledge production. This kind of epistemology analyzes generic internet 'users' interacting with other generic 'users' online, rather than talking about how users' social identities and locations in oppressive systems shape online knowledge production. This book shows the value of a socially situated approach—one that draws on feminist epistemology, anti-racist epistemology, queer epistemology and other approaches that analyze the effects of power and inequality on knowledge production and dissemination.

---

[*] Professor of Philosophy, Hobart & William Smith Colleges and Visiting Associate Professor, the African Centre for Epistemology and Philosophy of Science, University of Johannesburg. Her main lines of research are ethics, philosophy of science, feminist epistemology and social epistemology. Recent publications: Frost-Arnold, K. (2021). "The Epistemic Dangers of Context Collapse Online." In *Applied Epistemology*, (ed.) J. Lackey. New York: Oxford University Press; Frost-Arnold, K. (2020). "Trust and Epistemic Responsibility." In *The Routledge Handbook of Trust*, (ed.) J. Simon. New York: Routledge.
frost-arnold@hws.edu

Who Should We Be Online? applies several existing socially situated epistemological frameworks to the internet. Chapter one provides an introduction to feminist accounts of objectivity, veritistic social epistemology, epistemologies of ignorance, virtue epistemology, and epistemic injustice. I show how these frameworks fit together to provide mutually reinforcing evaluative tools for determining the epistemic merits and flaws of features of the internet and our actions as online agents. The rest of the book applies these frameworks to epistemic challenges raised by the internet. Each chapter focuses on one or two epistemically significant personas that populate the internet: moderators, imposters, tricksters, fakers, and lurkers. Chapter 2 investigates the epistemology of online content moderation, arguing that current corporate practices promote epistemic injustice and exploit workers in traumatizing ways. In Chapter 3, I examine internet hoaxes. I argue for a crucial distinction between internet imposters who cause epistemic damage by violating norms of authenticity and internet tricksters who violate these norms in acts of resistance that encourage epistemically beneficial trust in the oppressed. Chapter 4 addresses fake news. I show how racism shapes online disinformation and how disinformation fuels racism. I argue that feminist accounts of objectivity can provide tools for platforms to avoid what I call 'a flight to neutrality' that prevents them from accepting responsibility for their role in the fake news problem. In Chapter 5, I analyze how social media can play a powerful role in educating people about their own privileges and prejudices. I focus on the epistemic virtues and vices of lurkers, who are people who spend time in online epistemic communities without directly participating in them. I develop a virtue epistemology that helps us discern when to engage in a conversation, when to be quiet and lurk, and how to avoid hijacking online spaces for marginalized people. The research ethics appendix lays out several key ethical issues facing philosophers studying the internet, including privacy, protection of the researcher, and how to avoid epistemic exploitation of users.

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 157-159
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico)

Review of FROST-ARNOLD, K. (2023) *Who Should We Be Online? A Social Epistemology for the Internet.* New York: Oxford University Press (2023)

Las posibilidades que internet ofrece para aprender sobre realidades ajenas a la propia son amplias y están al alcance de cualquiera. Actualmente, un usuario poco familiarizado con el activismo antirracista o la lucha por los derechos LGBT+ puede conocer los entresijos de estos movimientos sociales a través de usuarios implicados en ellos, aunque estos no publiquen necesariamente con fines pedagógicos. Sin duda, y tal y como plantea Karen Frost-Arnold en su libro *Who Should We Be Online? A Social Epistemology for the Internet*, internet constituye un medio poderoso para desaprender la ignorancia construida por un determinado entorno social. Sin embargo, este mismo medio también puede llegar a menoscabar las buenas intenciones de los usuarios y perpetuar la ignorancia epistemológica de las personas.

En este libro publicado en 2023 por la editorial Oxford University Press, Karen Frost-Arnold, profesora en Hobart and William Smith Colleges, ofrece un relato interdisciplinar de los potenciales y los desafíos de internet basado en un marco de referencia feminista interseccional. La autora aglutina reflexiones teóricas de distintas disciplinas que abordan el estudio de internet, casos de estudio concretos y claves de aprendizaje no intrusivo que poner en práctica. En conjunto, se proporciona al lector una visión panorámica de los debates actuales sobre la epistemología social de internet y las dinámicas de raza, clase, género u orientación sexual que la moldean. El libro comienza con un capítulo introductorio que sienta las bases del marco que se utiliza a lo largo de la narración: el enfoque FOVIVI, que significa relatos Feministas sobre la Objetividad, el Veritismo, la Ignorancia, las Virtudes y la Injusticia. Los siguientes capítulos se articulan alrededor de cuatro personajes del entorno digital: los moderadores de contenido, los impostores, los creadores de noticias falsas y, por último, los *lurkers.* Con los tres primeros, la autora expone las amenazas y sesgos del mundo online, y con los *lurkers*, arroja luz sobre los beneficios epistémicos y limitaciones de una práctica, la de observar sin participar activamente, con la que aprender y viajar a "los mundos" de otras personas menos privilegiadas (Lugones, 2003).

A través de los moderadores, el primer personaje, la autora evidencia el primer filtro de sesgo por el que pasan las publicaciones que consumen los usuarios de las redes sociales.

A pesar de la creencia popular, estas plataformas no son espacios sin normas en los que actuar con total impunidad, sino espacios altamente regulados por moderadores, que adoptan el papel de *gatekeepers*. Estos trabajadores dan forma al flujo de información y, al mismo tiempo, los valores y las normas de la comunidad online que vigilan. Sin embargo, suelen tomar sus decisiones de sanción en cuestión de segundos, lo que termina por favorecer a grupos dominantes y perjudicar a comunidades marginadas. Los moderadores se ven sujetos a condiciones laborales injustas, lo que redunda en más ignorancia epistemológica para los usuarios, que no pueden evitar exponerse a contenido sesgado.

Con los impostores y los creadores de noticias falsas, la autora analiza a los usuarios infiltrados en comunidades y los perjuicios que conlleva su comportamiento. Sus engaños, por un lado, vulneran el conocimiento de la verdad, y por otro, diezman la confianza de la audiencia en las personas pertenecientes a grupos minoritarios. Los impostores, concretamente, traicionan las expectativas de autenticidad que se espera en este tipo de entornos. Antes de ser descubiertos, tienden a alimentar estereotipos equivocados sobre las comunidades a las que dicen pertenecer, si bien este daño no es manifiesto. Después de que su engaño salga a la luz, envenenan la comunidad epistémica a la que han fingido pertenecer con desconfianza. El público dudará de la honradez del resto de miembros y se mostrará reticente a escuchar lo que tienen que decir, lo que silenciará sus voces del discurso dominante aún más. Esta problemática va de la mano de las noticias falsas y la "ignorancia blanca" que promueven, fruto de una objetividad malentendida por parte de las compañías de redes sociales. La autora enfatiza cómo estas falsedades prolongan narrativas racistas, que pueden llegar a saturar la memoria colectiva durante muchos años.

El último capítulo se antepone a los anteriores con una visión optimista, pero realista, de las oportunidades que ofrece internet para aprender sobre epistemología social. La autora expone su argumento a través de la figura de los *lurkers.* Estos usuarios han quedado apartados del imaginario colectivo debido a la narrativa de "la cultura participativa de internet" y las numerosas investigaciones centradas en los usuarios que publican y diseminan contenido. Sin embargo, la evidencia empírica es clara: la gran mayoría de usuarios de internet prefiere observar lo que los otros publican sin participar en el flujo de información activamente (Nielsen, 2006). La figura del *lurker* ha arrastrado tradicionalmente una connotación negativa, quizás por las expectativas participativas que se tenía de internet al inicio, y frecuentemente se les ha tratado como miembros de segunda clase de las comunidades online, que contribuyen poco o muy poco al valor de estas. También ha habido algunos académicos que han reflexionado sobre el uso pasivo de internet como una forma legítima de participación. Pues, si bien los *lurkers* son pasivos, aprenden activamente sobre la comunidad que observan y su complejidad y dinámicas.

Así bien, el *lurking* es relevante porque permite al usuario aprender de otras comunidades y encontrar respuestas a sus preguntas sin involucrarse directamente en ellas y, lo que es igual de importante, sin cargar a las personas de dichas comunidades con el peso de tener que enseñarles. Al mismo tiempo, este comportamiento también evita que las personas recién llegadas dañen desde su desconocimiento a los miembros de una comunidad. Sin embargo, como señala la autora, ni siquiera esta figura está exenta de limitaciones epistémicas y éticas. Antes de llevar a cabo el *lurking,* el usuario deberá cerciorarse de si el espacio digital en cuestión es apto para la observación. Es decir, comprobará si es una esfera privada donde

las personas prefieren no ser observadas. Además, el proceso de aprendizaje probablemente requiera la implicación del observador en algún momento para que el viaje a estos mundos desconocidos implique un cambio real en la persona.

En definitiva, *Who Should We Be Online? A Social Epistemology for the Internet*, ayuda a comprender, a través de cuatro figuras relevantes del espacio digital, los vicios y virtudes de este espacio, cada vez más presente en el día a día de los ciudadanos. En internet, los usuarios pueden elegir aislarse en una cámara de eco que refuerce su visión sobre el mundo. Si el usuario toma este camino, está eligiendo voluntariamente cerrar los ojos a las oportunidades que el mundo online ofrece para aprender sobre los privilegios epistémicos de ciertas realidades y los prejuicios de su conocimiento. El camino alternativo, el del aprendizaje, conduce a resultados más positivos y queda trazado efectivamente por Karen Frost-Arnold a lo largo de su libro. Las investigaciones futuras disponen ahora de una recopilación interdisciplinar de los estudios de internet sobre la producción de conocimiento online, que facilitará enormemente los debates interdisciplinares que requieren la complejidad de los desafíos de internet. Los lectores, por su parte, cuentan con un mapa que refleja el entramado de sesgos, riesgos y beneficios con los que se encontrarán en las comunidades de internet mientras aprenden.

Beatriz Jordá
(Universidad Carlos III de Madrid y Saint Louis University - Madrid Campus)
jordabeatriz@gmail.com

Referencias

Lugones, M. (2003). *Pilgrimages/peregrinajes: Theorizing coalition against multiple oppressions*. New York: Rowman & Littlefield Publishers.

Nielsen, J. (2006). *The 90-9-1 rule for participation inequality in social media and online communities.* Recuperado de https://www.nngroup.com/articles/participation-inequality/

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 161-168
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.610871

What about my true beliefs? On the construction of our collective memory online

¿Y mis creencias verdaderas? Sobre la construcción de nuestra memoria colectiva en línea

*LOLA MEDINA VIZUETE**

Abstract: By applying Mills' notion of 'collective memory', Frost-Arnold argues that an *excessive* number of *false* beliefs online (fake news) can condition the memory that we share as a collective. Here I suggest, following Mills' original characterization of 'ignorance', that the construction and maintenance of our collective memory is also vulnerable to a partial *lack* of or total *absence* of *true* beliefs online. I suggest we must investigate these beliefs attending to two issues: firstly, instances of knowledge that are underrepresented, and secondly, non-propositional forms of knowledge. The first problem is addressed in section 1, where I explore different ways in which some beliefs might not reach the online sphere, due to their minoritarian status. The second problem is the focus of section 2, which entails the consideration of non-dominant forms of knowledge: knowledge-how and knowledge by acquaintance.
Keywords: collective memory, fake news, knowledge-how, knowledge by acquaintance, epistemology of internet.

Resumen: Aplicando la noción de 'memoria colectiva' de Mills, Frost-Arnold argumenta que un *exceso* de creencias *falsas* en línea (fake news) puede condicionar la memoria que compartimos como colectivo. Aquí sugiero, siguiendo la caracterización original de 'ignorancia' de Mills, que la construcción y mantenimiento de nuestra memoria colectiva también es vulnerable a cierta *falta* o *ausencia* total de creencias *verdaderas* en línea. Propongo que debemos investigar estas creencias atendiendo a dos cuestiones: en primer lugar, a instancias de conocimiento que están subrepresentadas y, en segundo lugar, a formas no proposicionales de conocimiento. El primer problema se aborda en la sección 1, donde exploro diferentes formas en las que algunas creencias pueden no alcanzar el ámbito digital debido a su estatus minoritario. El segundo problema es el foco de atención en la sección 2, donde se consideran formas no dominantes de conocimiento: el saber-cómo y el conocimiento por familiaridad.
Palabras clave: Memoria colectiva, fake news, saber-cómo, conocimiento por familiaridad, epistemología de internet.

Recibido: 04/04/2024. Aceptado: 24/06/2024.

* PhD candidate at the University of Sevilla (lmvizuete@us.es). Her research focuses on social epistemology, with special attention to epistemic injustice, feminist and gender issues. This work was supported by [NANORIN] (The Nature and Normativity of Inquiry), a project funded by the Spanish Government under Grant [PID2021- 123938NB-I00]; [METAPRODES] (Meta-actitudes, desacuerdos profundos y progreso moral) a project funded by the Spanish Government under Grant [PID2021-124152NB-I00]; and [digi_morals] (Moral disagreements on the digital sphere. Interactive dynamics, micro-mechanisms and cultural markers), a project funded by BBVA Foundation.

Who should we be online? (2023) is an exciting and normative proposal about our roles and duties as epistemic agents in an environment that is no longer new: the Internet. One of the most original contributions of the book is the update of Charles Mills' epistemology to the digital environment (section 4.4. Fake News and White Ignorance) and, specifically, Frost-Arnold's use of Mills' notion of 'collective memory' to analyze the phenomenon of fake news. Along this section, Frost-Arnold convincingly defends that to understand the creation and spreading of fake news in our epistemic environments we must comprehend the implications of white racism and white domination in our daily epistemic practices. In order to defend this claim, she aligns herself with two crucial notions. Firstly, what we are is collectively construed and collected. This means that our collective memory is a social effort to maintain and produce our group identity and history. Secondly, racist beliefs and practices (what Mills labels as white ignorance) greatly determine what or who we remember, appreciate, celebrate, or recognize socially. It follows from this that the construction and maintenance of our collective memory would be shaped and influenced by white ignorance as long as our beliefs and collective resources are shaped and affected by racist commitments. The content available online and the ways we access it (both actively and passively) are crucial ways in which our collective memory and practices are molded and kept alive. Search engines, social media content, and the lack of active exercise of moderation online are some ways in which prejudicial beliefs spread and settle. Fake news, along these lines, may be understood as an online manifestation of white ignorance that determines our collective memory and testimonial practices online.

Frost-Arnold's analysis thus targets a particular kind of false beliefs: those that generate a type of miscognition that distorts our collective memory. Her rationale crucially focuses, then, on how these false beliefs are easily replicated and how the exposure to this online content effortlessly expands, not only in cases with spurious intentions but even when the intention of the content is to debunk the original false belief (125). Hers is, in other words, a concern about how an *excessive* number of false beliefs can condition the memory that we share as a collective.

In this comment, I would like to emphasize that Frost-Arnold pays less attention to the *second* way in which Charles Mills characterizes white ignorance, as the possibility of a miscognition derived from the *absence* of true beliefs, instead of merely originating from the presence of false beliefs (Mills, 2007, 16). My aim, following Mills' original characterization of ignorance, is to put some pressure on Frost-Arnold innovative proposal by suggesting that the construction and maintenance of our collective memory is also vulnerable to a partial lack of or total absence of true beliefs online. Therefore, I want to claim that, together with the worry about the spread of falsehoods online (such as fake news and other misrepresentations), we should also pay due attention to those lacunae, namely: beliefs that, even if they constitute common knowledge in the offline world, are underrepresented online, or only shared by a minority of users[1].

1 Some might raise doubts about the mere possibility of a collective memory online given the existence of extreme personalization techniques. For my purposes it would be enough to state that I find Frost-Arnold's considerations in this regard persuasive (Frost-Arnold, 2022, 145). Therefore, if the reader is not convinced by the possibility of a collective memory online there are still good arguments to, at least, grant the idea of a 'perceived collective memory'.

To argue for the impact that the absence of true beliefs has on collective memory, I suggest we must investigate these beliefs attending to two issues: firstly, instances of knowledge that are underrepresented, and secondly, non-propositional forms of knowledge. The first problem is addressed in section 1, where I explore different ways in which some beliefs might not reach the online sphere, due to their minoritarian status. The second problem is my focus on section 2, which entails the consideration of non-dominant forms of knowledge: knowledge-how and knowledge by acquaintance.

1. Minorities and their beliefs

It is clear, from Frost-Arnold's analysis, that excessive speakers sharing and disseminating racist, sexist, ableist, or any kind of false beliefs in the online environment affect the construction and maintenance of our collective memory. Here I would argue that it is also relevant to pay attention to how such memory is construed when some realities or voices have little to no presence in the online environment.

To be fair, Frost-Arnold's analysis does not ignore the dangers entailed by the scarcer contributions of certain communities to the Internet. On the contrary, she convincingly deals with these affairs. I would press her analysis because she is foremost concerned (and rightly so) about communities that are excluded from the digital space *due to* epistemic or social injustices. In this way, she makes the case for the objective difficulties that certain communities or agents can suffer online when trying to equally participate or gain due credibility in the production and dissemination of knowledge. My worry is that she does not engage with the possibility of a *defective* presence of true beliefs in the online sphere *for other reasons* than unjust marginalization or injustices more generally. The lower presence in the online sphere of some instances of knowledge about certain issues might be the result of communities that simply comprehend fewer speakers or scarcer members[2]. This means that true beliefs of minoritarian groups, by the simple fact of them being a minority, either offline or online, might be at risk of not reaching the online sphere and consequently, not participating in the configuration of the collective memory.

A first consideration here is that there are several ways in which it is possible to refer to a minoritarian presence of certain groups (and/or their beliefs) in the online sphere. Consider the following possibilities:

- The community's online presence and their beliefs, knowledge, and understandings are *accurately represented* in the digital environment regarding the existing number of members. In this case, the community has few members, given the amount of people that qualify to be a part of it. Think, for example, about beliefs shared by communities of patients of rare diseases, societies that share a language that has very few speakers, or social groups that are really reduced in numbers, such as some indigenous populations. For such cases, an accurate representation of the members of these groups online, considering the true-life members, could result in a low quantity of online content from this group more generally.

2 This minoritarian status might derive from a quantitative reality (few members) or from qualitative circumstances (lack of resources, interest, political commitments…). More on this below.

- The community is not necessarily composed of a reduced number of members, but *their online presence* is substantially smaller, specifically due to limited access to the resources that enable participation in the online dimension. It is barely controversial that, to access the Internet, it is necessary to enjoy certain material and non-material resources (privileges?) such as electronic devices, Internet bandwidth, digital literacy…[3] Lacking any of these resources may result in a significant reduction of the participation of members of some communities in the online sphere, which are nonetheless significant in number in offline spaces. A simple example of this common scenario is the disproportionately low presence of the elderly population in some social media contexts compared to the increasing numbers of aging populations in Western societies. Rural or impoverished areas, where the population has little access to internet coverage or public resources for internet usage are scarce, could also confront similar disproportion in the presence of their members online. Or, in a more controversial picture, young people and infants have limited access to several online spaces where their beliefs and knowledge could be relevant (discussion of children's rights, city management, educational content…).
- Communities that might not be reduced in factual numbers but *do not wish to engage* with the digital environments for various reasons. In this case, the absence of possible true beliefs from some online communities would not have originated from a low number of participants, but from a decision not to be part of the digital sphere. In different ways, concerns about the hypervigilance of digital devices or data transfer online could motivate such a refrain from the digital world[4]. For instance, activist groups that are committed to protecting the privacy of individuals by not using certain capitalized online spaces or that aspire to reduce their ecological impact by choosing to disengage as much as possible from specific domains or from the Internet altogether. Somewhat more problematic motivations to disengage from the digital could also originate in commitments to conspiracy theories or cults that choose to avoid digital environments.[5]

Regardless of which of these reasons are grounds for lower participation of certain communities in the digital environment, there is a risk that some true beliefs or knowledge possessed by these minoritarian communities is being neglected online. Following Mills and Frost-Arnold, this could impact the construction and maintenance of collective memory, since there would be a deficit of certain voices online, that could differently shape this memory.

---

3 Some may worry that these are precisely the circumstances that ground the epistemic injustices and forms of oppression that Frost-Arnold worries about, and that, therefore, the minoritarian communities referenced here are the very ones targeted by her book. Settling the question about which digital resources should be granted in a just society clearly exceeds the scope of this comment. For the present discussion, it is enough to state that I believe in the possibility of a lack of resources or capacities for participation in the online sphere that is not grounded in unjust circumstances. For example, lacking the digital literacy that digital natives possess can result from a disinterest in new technologies.

4 It could be argued that completely refraining from the digital environment is impossible since there is no longer such a division between online and offline spaces. A more fruitful way to engage in this debate could be to explore if this disengagement from the digital sphere is *de facto* a possibility or if the capacity to disengage from it actually resides in some kind of privilege or advantage that just some people enjoy. I suspect the latter.

5 In this case it is certainly more difficult to analyze the real possibility that this communities could contribute with true beliefs or knowledge to the shared pool of knowledge.

At this point, a second consideration is in order. Some might worry that there are no *epistemic* reasons to argue for such a presence of minority groups in the online sphere. Although several *political* or *moral* considerations could clearly legitimize a vindication of their presence online, it might not be straightforward *why* these minority groups should be equally present online, in terms of epistemic reasons. However, these concerns should disappear when we consider how the contribution of these minoritarian groups entails potential *epistemic* benefits (or even privileges). In fact, there are good reasons to think that individuals belonging to minoritarian communities might enjoy better epistemic locations regarding certain realities (Du Bois, 1897; Haraway, 1988; Harding, 1992; Medina, 2013). To use some of the aforementioned examples, consider how people living with rare diseases are in better positions than those who don't suffer from any severe condition to know general and specific claims about the health system, due to their need to understand it and their acquaintance with it. In the same way, older people are potentially epistemically better suited to understand the revolution of technologies for day-to-day activities, precisely because they have first-hand knowledge of a time when digital technologies did not exist.

Once one agrees with the potential epistemic benefit of minority groups contributing to the general pool of knowledge, it becomes also relevant that communities that are not unjustly marginalized (at least explicitly) can fail to contribute to the online domain with their true beliefs. In the first place, because potentially relevant true beliefs can be collectively neglected. Added to that, the inquiries from majoritarian groups about issues they ignore would be hard to settle. This could be the case if we consider how lower online participation of minoritarian groups amounts to a deficit in the quantity of available content on topics that only they could produce online. Consequently, breaking through issues that majoritarian communities ignore (and minority communities might shed light on) would entail high epistemic labor and cost for those who are concerned about settling questions about said issues. Put simply, digital technologies such as search engines would be able to easily provide multiple sources and content for topics that are widely shared (the results of the US elections in 2020, for example) but few results for themes for which little content is created and shared (what are the common symptoms of menopause, for instance). As a result, there could be a deficit in the collective memory since relevant contributions from minoritarian groups would not get to model it.

2. Non-propositional knowledge

The same way the collective memory could be affected by the absence of true beliefs due to minorities not contributing in equal numbers to the shared pool of knowledge, this collective memory might be impacted if we dismiss some types of knowledge, such as know-how or acquaintance, in favor of others, such as propositional knowledge (see Shotwell, 2017). Of course, there is still space for controversy regarding the possibility of irreducibly non-propositional modes of knowledge[6]. But even if some reductions were manageable, non-propositional forms of knowledge would still instantiate peculiarities that we may want to

6 The debate on the possibility of non-propositional forms of knowledge is still open. A good revision of the state of the art for 'knowledge-how' can be found in Navarro (2021). A thoughtful revision of 'Contemporary Views on Acquaintance' and their criticisms can be found in Hasan & Fumerton (2020) and in Ducan (2021).

preserve and promote. Consequently, a misrepresentation of these non-propositional forms of knowledge may affect the creation and maintenance of the collective memory online. My aim in this section is to motivate these considerations.

A fruitful way for many philosophers to argue for the possibility of non-propositional forms of knowledge has been to defend some distinctive features in these that are not present for knowledge-that (Navarro, 2021). I want to suggest that these unique features that tell knowledge-how and knowledge by acquaintance apart from knowledge-that might affect how beliefs are shared, produced, and questioned online. Furthermore, I claim that only attending to or prioritizing propositional knowledge over other forms of knowledge shapes our collective memory in defective ways.

To make the case for such a claim, consider some of the unique characteristics attributed to knowledge-how. Contrary to knowledge-that, it is persuasively argued (Hawley, 2010; Poston, 2016) that knowledge-how is resistant to testimonial transmission. This implies, for example, that it is not possible to convey how to pilot a plane just by communicating some propositional truths about the practice of flying an aircraft, instead, to know how to pilot, these truths must be connected to the action of flying. In the same way, knowledge-that is widely considered to be an all-or-nothing state (Drestke, 1981)[7]. Either you know that today is Monday, or you don't. This is arguably not the case for knowing-how (Bengson and Moffett, 2011; Sgaravatti & Zardini, 2008; Pavese, 2017). It is possible to know, for instance, how to play football in various degrees; as an amateur player that meets their friends on the weekends, or as a devoted professional. Similar arguments are also in place for the specific features of knowledge by acquaintance.

In the case of acquaintance, these unique features are even clearer since to be acquainted with anything is to have direct awareness of it (Russell, 1911, 1912). Some understand this direct awareness narrowly, namely, as a completely unmediated relation (Fumerton, 1995; BonJour, 2001). Thus, there is just the possibility of being acquainted with one's states of mind (phenomenal properties such as colors, smells, pain, itchiness…). Others, however, understand this directness in broader terms and argue that one can be acquainted with physical objects or people, and to know them (Brewer, 2011; Tye, 2009). For what is worth here, to be acquainted with something, someone, or somewhere (a color, a relative, a city…) one does not need to hold true propositions about them, but just enjoy a direct awareness of them, in either of the preferred senses. Additionally, this direct awareness ensures that knowledge by acquaintance comprehends distinctive attributes. For instance, it has been argued that this type of knowledge is, first, especially complete and, secondly, distinctively secure (Russell, 1912). In this way by being acquainted with pain, for instance, one does not only know about the pain completely but also has some knowledge that is indubitable.

Considerations as the ones highlighted about the unique features[8] of non-propositional forms of knowledge are crucial to understanding how some types of knowledge are present in the online environment or not, and if they are, how they differently shape which modes

7 Sosa's notion of 'knowing full well' is an exception to this consideration (Sosa, 2011).

8 Several other distinctive features have been discussed in the literature on know-how (resistance to veritic intervening luck (Potson, 2009), resistance to environmental epistemic luck (Carter & Pritchard, 2015), resistance to epistemic defeaters, (Carter & Navarro, 2017). It remains debatable whether they could be relevant to the arguments defended here.

we share and treasure as a collective. Remarkably, paying attention to these specific issues about non-propositional knowledge unveils at least two concerns about them and the Internet. First, non-propositional modes of knowledge may face difficulties in entering the digital space due to their unique features. It is possible to make the case for an absence of certain kinds of knowledge that, due to their unique characteristics, would not enter the online domain. There is a risk, for instance, of losing knowledge about how to produce textiles in artisanal ways or how to cure some diseases with ancestral techniques precisely because the digital sphere is not a good candidate for the preservation of modes of knowledge that are resistant to propositional ways of conservation and transmission. There are also strong difficulties in arguing for the acquisition of any knowledge by acquaintance in the online dimension, besides knowledge about the digital environment itself (the online features, the technological affordances, the dynamics of social media…). But even if agreed that this kind of knowledge can be accommodated in the online sphere (think about tutorials, simulators to teach professionals, augmented reality, media archives…), which is a claim that is subject to dispute, there is a second worry that we should account for. This is that propositional accounts of knowledge might be prioritized over non-propositional ones because they are a better candidate for a canonical mode of transmission online: testimonial transmission. Think, for example, how easier it is to acquire some propositional truths about Athens online (e.g. it is the capital of Greece, the Parthenon is there…) compared to the acquisition of any acquaintance with the classical beauty of their monuments or with the high temperatures endured during summers.

Therefore, there are good reasons to attend to non-propositional forms of knowledge in the online sphere, since there is a risk of losing or neglecting certain kinds of knowledge online that could enrich the shared pool of knowledge. Consequently, there is a risk that our collective memory might become defectively construed and maintained, due to an absence of non-propositional forms of knowledge.

3. Conclusion

The described ways in which the absence of true beliefs online might impact our collective memory are theoretically differentiated here to better understand how lacking true beliefs online could affect our collective memory. Nevertheless, these descriptions are not sealed from each other in the online space. On the contrary, the issues outlined (minority condition and non-propositional kinds of knowledge) can condition particular realities at the same time. This can be the case, for example, of minorities that cannot significantly contribute with their knowledge-how to the internet. The loss or absence of their relevant true beliefs in the online space would surely affect our collective memory.

References

Bengson, J., & Moffett, M. A. (2011). Nonpropositional Intellectualism. In J. Bengson & Moffett, Marc A (Eds.), *Knowing how: Essays on knowledge, mind, and action,* 161-195. Oxford University Press.

BonJour, L. (2001) Toward a Defense of Empirical Foundationalism. In Michael Raymond DePaul (ed.), *Resurrecting Old-Fashioned Foundationalism* (pp. 21-38). Lanham: Rowman and Littlefield.

Brewer, B. (2011). *Perception and its Objects*. Oxford: Oxford University Press.

Carter, J. A., & Navarro, J. (2017). The Defeasibility of knowledge-how. *Philosophy and Phenomenological Research, 95*(3), 662-685. https://doi.org/10.1111/phpr.12441

Carter, J. A., & Pritchard, D. (2015). Knowledge-How and Epistemic Luck. *Noûs, 49*(3), 440-453.

Dretske, F. (1981). The pragmatic dimension of knowledge. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 40*(3), 363-378.

Du Bois, B.W.E. (1897), The Strivings of the Negro People, *The Atlantic Monthly*, August: 194–197.

Duncan, M. (2021). Acquaintance. *Philosophy Compass, 16*(3).

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

Fumerton, R. (1995). *Metaepistemology and Skepticism*. Lanham: Rowman and Littlefield.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575. https://doi.org/10.2307/3178066

Harding, S. (1992). Rethinking Standpoint Epistemology: What is «Strong Objectivity?». *The Centennial Review*, *36*(3), 437-470.

Hasan, A. & Fumerton, R. (2020, Spring) Knowledge by Acquaintance vs. Description. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/knowledge-acquaindescrip/>.

Hawley, K. (2010). Testimony and knowing how. *Studies in History and Philosophy of Science Part A*, *41*(4), 397-404.

Mills, C. (2007). "White ignorance". In Sullivan, S. & Tuana, N. (Eds) Race and epistemologies of ignorance, (pp. 26-31). State University of New York Press.

Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and the social imagination.* Oxford University Press.

Navarro, J. (2021). Knowing how to. *Routledge Encyclopedia of Philosophy*. URL = https://www.rep.routledge.com/articles/thematic/knowing-how-to/v-2#

Pavese, C. (2017). Know-how and gradability. *Philosophical Review, 126*(3), 345-383.

Poston, T. (2009). Know how to be Gettiered? *Philosophy and Phenomenological Research, 79*(3), 743-747. https://doi.org/10.1111/j.1933-1592.2009.00301.x

Poston, T. (2016). Know how to transmit knowledge? *Noûs, 50*(4), 865-878

Russell, B. (1911). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11*, 108–128.

Russell, B. (1912). *The problems of philosophy*. London, UK: Thornton Butterworth Limited

Shotwell, A. (2017). Forms of Knowing and Epistemic Resources. In Kidd, I. J., Medina, J., & Pohlhaus Jr, G. (Eds.). *The Routledge handbook of epistemic injustice*. Routledge, Taylor & Francis Group, 79-88

Sgaravatti, D., & Zardini, E. (2008). Knowing How To Establish Intellectualism. *Grazer Philosophische Studien*, *77*(1), 217-261. https://doi.org/10.1163/18756735-90000849

Sosa, E. (2011). *Knowing Full Well*. Princeton: Princeton University Press.

Tye, M. 2009. *Consciousness Revisited.* Cambridge: MIT Press

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 169-178
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.612021

On testimonial justice online. Nuancing Karen Frost-Arnold's optimistic virtue epistemology

Sobre la justicia testimonial online. Matizando el optimismo de Karen Frost-Arnold acerca de la epistemología de la virtud

GONZALO VELASCO ARIAS[*]

Abstract: In Who Should We Be Online, Karen Frost-Arnold advocates an approach to epistemic virtues that resists pessimism about the possibility of our online epistemic agency being responsible and socially just. On the basis of a veritist epistemology, her proposal overcomes both responsibilist individualism and the socio-structural critique that delegates all responsibility to institutional transformations. The author identifies in online lurking an activity unique to online epistemic agency that can provide exposure to messages from people discriminated against by epistemic injustices. For Frost-Arnold, moreover, this implies the possibility of the lurker experiencing epistemic frictions that will favour a more reliable willingness to be fair in giving credit to the testimonies of those discriminated against. In this note I will qualify this optimistic stance, arguing

Resumen: En *Who Should We Be Online*, Karen Frost-Arnold defiende una cierta aproximación a las virtudes epistémicas que resista al pesimismo acerca de la posibilidad de que nuestra agencia epistémica online sea responsable y justa. Sobre la base de una epistemología veritista, su propuesta supera tanto el individualismo responsabilista como la crítica socio-estructural que delega toda responsabilidad en transformaciones institucionales. La autora identifica en el online lurking (fisgoneo online) una actividad exclusiva de la agencia epistémica online capaz de proporcionar una exposición a mensajes de personas discriminadas por injusticias epistémicas. Para Frost-Arnold, a su vez, esto implica la posibilidad de que el lurker experimente fricciones epistémicas que favorecerán una disposición más fiable a ser justos a la hora de dar crédito a los testimonios de personas discriminadas. En esta nota matizaré esta postura

Recibido: 12/04/2024. Aceptado: 27/06/2024.

* Gonzalo Velasco Arias es Profesor de Filosofía en la Universidad Carlos III de Madrid. Sus principales líneas de investigación son (1) la aproximación filosófica a las emociones políticas, (2) las injusticias epistémicas en el espacio público digital y (3) el análisis de las manifestaciones subjetivas de las injusticias estructurales. En relación a estos temas, recientemente ha publicado el artículo "Arrogancia, desconfianza y desafección. Una aproximación desde la epistemología de la virtud", *Isegoría. Revista de Filosofía moral y política*, N.º 70, enero-junio, 2024, https://doi.org/10.3989/isegoria.2024.70.1348; el ensayo *Pensar la polarización*, Gedisa, Barcelona, 2023; así como la coedición junto a A. Gómez Ramos del volumen colectivo *Atlas político de emociones*, Trotta, Madrid, 2024. Es miembro de los proyectos de investigación "El vínculo y su contrario. Desafección, mediaciones y representación política (VI_CO)" (PID2021-124954NB-I00), "Desacuerdos morales en la esfera digital. Dinámicas interactivas, micromecanismos y marcadores culturales (digi_morals)", Fundación BBVA. gvelasco@hum.uc3m.es.

the epistemic individualism that underlies it. I will point to a group virtue model as a possible solution.
Keywords: epistemology of virtue, online testimony, deference, lurkers, trust, humility.

optimista, alegando el individualismo epistémico que subyace. Apuntaré a un modelo de virtudes grupales como posible solución.
Palabras clave: epistemología de la virtud, testimonio online, deferencia, fisgones, confianza, humildad.

1. Introduction

Online epistemic virtue is the focus of Karen Frost-Arnold's enquiry in Who Should We Be Online (Frost-Arnold, 2023). The research objectives of the book can be synthesised in the attempt to outline the conditions under which responsible epistemic agency can take place in the hybrid media space (Chadwick, 2017), including in this consideration of responsibility the dimension of epistemic justice. Far from indulging in a naïve conception of virtue, her enquiry embeds an evaluative reflection on the epistemology of virtue that allows her to refine a model applicable to the analysis of online agency. In my view, this takes the form of two issues, which I outline as an introduction to the rest of my argument.

First, Frost Arnold eschews the nostalgic metanarrative that argues that the epistemic and informational problems of our hybrid media system can be solved by recovering or enhancing classical virtues. Instead, she advocates an epistemology of situated virtue that allows for a non-ideal approach to the normative issues that arise from our online epistemic agency. In addition, these situated virtues are framed in a systems-oriented social epistemology that considers that what matters is not identifying which virtues are useful for the risks inherent in online epistemic agency (Heersmink, 2018; Vallor, 2016) (Driver, 2001, p. 68), but identifying which vices and which virtues are fostered by the epistemic structure of the network. This, again, entails a response to the exceptionalist narrative. This view redresses the motivationism and doxastic voluntarism behind the responsibilism that is present in the social discourse and in many approaches that still understand education and training in offline epistemic virtues as the only way to improve epistemic behaviours and to mitigate the intrinsic risks of online communication (Heersmink, 2018, 10).

Secondly, the combination of virtue epistemology with a certain systems-oriented social epistemology advocated by Frost-Arnold enables to overcome the dichotomy between individual responsibilism and socio-structural critique (Anderson, 2012). Frost Arnold's work is an attempt to safeguard agency without yielding to naïve faith in individual virtue or to the maximalism of socio-structural critique.

Despite this sophisticated version of epistemic virtue, I believe that Frost-Arnold's normative proposals for epistemic agency online in relation to epistemic injustices remain grounded in a certain epistemic individualism. I will articulate my critique with reference to chapter 5, devoted to virtuous lurking. According to the author, virtuous lurking is a specific practice of online epistemic agency, which renders the exposure of privileged subjects to epistemic frictions and unlearning easier through the reception of testimonies issued by users belonging to objectively discriminated collectives, while avoiding the undesired effects of the interference of privileged agents in the conversation between discriminated agents.

Frost-Arnold argues that this allows for hopeful trust on the part of marginalised subjects. The possibility of these practices is thus a reason for optimism about the possibility of more responsible agency in relation to epistemic injustices online. In order to elaborate my account of Frost-Arnold's optimism, I first need to make some analytical clarifications about (1) approaches to online testimony, (2) the account of testimonial justice as a virtue, and (3) some specific phenomena of the digital epistemic environment that hinder responsible and justice-focused agency.

2. Testimony and online deference

In online or hybrid communicative contexts, restrictive definitions of testimony and deference are not useful. Narrow definitions of testimony are those that only accept as such speech acts in which the speaker intends to present evidence to someone about a matter known to be in dispute. Equally narrow is the view that limits testimony to speech acts in which the speaker claims that his statement is true, thereby committing the hearer to believe and trust him. Given the complex interactivity and ambiguity of speech acts in hybrid or online communicative environments, I believe it is appropriate to accept a permissive approach to testimony (E. Fricker, 1995), which with Sosa could be defined as "a statement of someone's thoughts and beliefs, which they might direct to the world at large and not one in particular" (Sosa, 1995, p. 219). Sullivan has argued that, on the Internet, specifically on social networks, although a statement can be reposted indefinitely without the original message being modified, the communicative meaning can change from the original (Sullivan, 2019). The consequence of this is that not even the informational basis of the testimony is preserved, which would be the condition required by approaches such as that of Lackey (Lackey, 2007). Therefore, strictly speaking, we could not speak of "testimonies" in digital environments.

Despite this, digital environments are in fact used as a source and means of expression of information and testimonies. Frost-Arnold provides a wide range of evidence of communicative situations in which discriminated groups use their online avatars to express and testify about the injustices they suffer. Regardless of the communicative acts they use to do so, Frost-Arnold sees this phenomenon as an extension of the production and circulation of true beliefs about situations of injustice, and as an opportunity for those who, because of their position of privilege, do not have access to these life experiences.

Although it is not a concept employed by Frost-Arnold, I believe that the problem of deference is implicit in reflection on the epistemic friction to which the Internet exposes privileged subjects. Heuristically, I think it is convenient to adopt a broad definition of deference, in the sense that "A defers to B on the question whether p if A believes that p (or not-p) merely because B believes that p (or not-p)" (Brinkmann, 2022, p. 267). This definition does not necessarily require either that there be intentional testimony or that the speaker be an expert in the domain of beliefs about which testimony is given, so it fits well with the notion of online testimony that I have just defined, and with the kind of listening that occurs to those who have testified to the suffering of epistemic injustice. Deference is a radical act of trust in which the listener assumes an epistemic risk that implies a strong

normative expectation in the trustworthiness and benevolence of the speaker. In other words, from the point of view of persuasive argumentation, it is irrational for a privileged subject to trust the testimony of a marginalised subject and eventually accept that this testimony forces him to change his own set of beliefs. I suggest thant when Frost-Arnold identifies online lurking as an opportunity for privileged agents to "unlearn" their prejudices and biases, he is not taking into account this excessive cost..

Deference in contexts of epistemic injustice, moreover, presents a difference with respect to modalities of moral or epistemic deference in more defined fields of knowledge or practice. The motivations for deference can be twofold: (a) the speaker's expression of his or her experience is convincing and persuades the listener to delegate the opinion to him or her, or (b) the listener recognises the testimonial and/or hermeneutic deficit to which the speaker is subject because of his or her group or social identity and, in order to compensate for this deficit, decides to trust him or her regardless of his or her agreement with his or her set of beliefs. In turn, testimony (in the broad sense referred to above) can be about the situation of epistemic injustice itself, or about a particular field of experience.

3. Is epistemic justice a virtue?

Virtue epistemology defines knowledge as a justified and non-accidental true belief that is a product of the agent's reliable epistemic competence. Applied to situations of testimonial injustice, it is virtuous "the hearer who reliably neutralize prejudice in her judgments of credibility". In other words, testimonial justice aspires to a state in which we do not too often find ourselves in the situation where we notice that we are committing unrectified epistemic injustices. In that case, we would be unreliable. For Fricker, self-corrective motivation is a necessary component for the reliability of the epistemically just agent. Although, in her responses to the criticisms raised by Shermann and Alcoff, she makes some concessions to a fallibilist version according to which it is possible to train and turn into habit the moments of apperception that motivate the correction of our epistemic vices (M. Fricker, 2010, p. 92). At the same time Fricker insists that such reliability can only be sustained by a hearer's motivation to do justice. Thus, although Fricker's own theory is an objection to doxastic voluntarism, there is a certain practical voluntarism.

The main criticisms of virtue epistemology come from "externalist" approaches such as situationism and socio-structural criticism. For socio-structuralism, the situations that impede the exercise of virtues are not contingent or trivial but regular and socially relevant, and have to do with structural power relations that shape our unconscious biases and prejudices. Hence, for Anderson, promoting individual virtues to redress structural epistemic injustice, while not wrong in itself, "plays a comparable role to the practice of individual charity in the context of massive structural poverty. Just as it would be better and more effective to redesign economic institutions so as to prevent mass poverty in the first place, it would be better to reconfigure epistemic institutions so as to prevent epistemic injustice from arising" (Anderson, 2012, p. 171).

The alternative to individualistic responsibilism, therefore, is often the institutional dimension. For Anderson, this does not exclude the promotion of individual virtue (Ander-

son, 2012, p. 166). However, from the point of view of the epistemology of virtue, the problem with the institutional solution is that if the application of virtue is conditional on institutional determinants linked to a concrete situation, then it no longer fulfills the condition of being cross-situationally consistent.

4. Arguments for pessimism

In my opinion, the main reasons for pessimism are the phenomena of information segregation, the spread of mistrust caused by the phenomenon of impostors and fakers, and also epistemic individualism that fosters "illusions of online understanding" (De Ridder, 2022). The question then is to what extent these virtuous practices have a normative force to overcome these difficulties. Since the first factor has been dealt with very abundantly by the specialised literature, I will focus on the last two.

4.1. The possibility of online imposture as a source of distrust.

Trustworthiness is a demand that the epistemic agent makes of the witness, to compensate for the risks associated with trusting. In trusting, the agent is vulnerable to intentional misinformation from the witness, and relies on the witness's knowledge of the issue to accredit the trust placed in him or her. Trust, therefore, requires normative expectations: in order to be trusted, the witness is expected to be both competent and benevolent (Dutilh Novaes, 2023; Levy, 2022). It is this second expectation, benevolence, that the online communicative environment systematically betrays because anonymity enables the phenomenon of the impostor. Anonymity makes it difficult to accredit the authenticity of the witness. Excluding other philosophical approaches to identity, I will restrict here the meaning of "authenticity" to consistency. The authenticity of an online user can then be encoded in his or her consistency, and this can be assessed as (i) the coincidence between the online and offline self, (ii) the consistency in action and self-presentation across platforms, and (iii) the consistency in online presentation over time (Frost-Arnold, 2023, p. 83). Those who present themselves consistently in the online debate are presumed to be reliable in ther intentions and behaviour[1].

Several authors have analyzed the distortions in trust generated by significant cases of impostors on various online platforms. Without going into the specificity of these cases, we can induce that the intrinsic possibility of online faking and imposture, the difficulty of verifying the authenticity of testimonies, and the very structure that fosters gaming and simulation, lead to the fact that, in conditions of vulnerability, mistrust in testimonies not accredited by authority or expertise is the default attitude (Frost-Arnold, 2023, pp. 77, 81). Online spaces are therefore hostile to the expression of testimonies of marginalisation and

1 As one reviewer has pointed out to me, users can be consistent in their inauthenticity. Indeed, consistency ii and iii do not imply benevolence or transparency of intentions at all. For this to happen, type 1 - consistency between the online and offline user - must also occur. And since, by definition, the current user model does not guarantee this requirement, I argue that anonymity is a structural source of distrust... unless progress is made in implementing some of the measures suggested by Veliz (2019), as I indicate below.

injustice. For this reason, it has been suggested that the major online social media platforms might modify their authenticity clauses, remove the right to anonymity or moderate it to pseudonyms whose authentic identity would be preserved by third party regulators who could apply progressive sanctions in case of abuse of the privileges of non-disclosure of authenticity (Véliz, 2019). However, it is also undoubtedly true that anonymity enables disruptive agencies that allow for the expression of testimony and political demands that under conditions of authenticity would be stifled by various forms of epistemic injustice. Other authors have argued for institutional markers of trustworthiness (Rini, 2017). However, these markers would increase epistemic gaps with those who do not have a prior trustworthy track record .

4.2. Online illusions of understanding.

From the point of view of research epistemology, although search engines are an undoubtedly agile tool for finding information, at the same time they feed "illusions of understanding" (De Ridder, 2022). Search engines, even more so in conversational versions such as ChatGPT, encourage confusing the mere connection of fragments of information with the process of comprehensive understanding (thus generating a false sense of self-sufficiency), pre-determine enquiry strategies with mechanisms for auto-completion of searches, impose arbitrary criteria for the evaluation of findings (order of appearance of information, popularity among users) and subject the agent to an infinite recursivity of information that makes it difficult to discern when sufficient evidence is available. This is the same effect caused by the availability and accessibility of expert opinion, as it can lead to a loss of track of our trustworthiness and distort our understanding of our own abilities (Fisher et al., 2015, p. 675).

The concept of "illusions of understanding" as applied to online enquiry raises two issues that need to be differentiated. The first (a.i) concerns perceptions of our own cognitive abilities and the fulfillment of our epistemic duties. By increasing our innate disposition to overestimate our abilities, it produces less reliable agents and, moreover, makes it difficult to correct this overestimation based on a certain amount of critical introspection. From a responsibilist point of view (i.b) "illusions of understanding" generate a false sense of fulfilling individual epistemic duties and thus foster less responsible agents through induced arrogance (Levy, 2019).

5. Lurking, optimism and the epistemic humility debate

To recapitulate, epistemic agency in hybrid online/offline environments tends to be more arrogant and individualistic due to ignorant overestimation of one's own capacities, and also less trusting due to a propensity for imposture that makes it difficult to assess the authenticity of testimony. Faced with the difficulty of identifying evaluation criteria that justify deference in testimonies of marginalised subjects, a first temptation would be to ask once again about the possibility of appealing to the intervention of the subjective intellectual virtues of the listener that would allow him to defer reliably. From the systems-oriented epistemology

approach employed by Frost-Arnold, what is interesting to know is whether the Internet is conducive to those dispositions that favour virtuous deference capable of compensating for the evaluative difficulty of online testimonies.

Karen Frost-Arnold identifies this possibility in the figure of the lurker. Lurkers are people who listen to, read or visualise expressions or communicative exchanges without participating in the communication themselves. The Internet, due to the guarantees of anonymity it provides, favours opportunities for this type of activity. The protection of anonymity would prevent defensive intuitive reactions that would arise in direct interpellation, and create the conditions for the benefits of "epistemic friction" or "world travelling" (Lugones), to mention some concepts that have been used to explain the possibility of the privileged learning about their own ignorance (Frost-Arnold, 2023, p. 174) by listening to the experiences and testimonies of discriminated groups. According to this view, lurking brings truthful benefits to the practitioner and avoids the unwanted side effects of other forms of well-meaning interaction of privileged allies with marginalised subjects. Above all, it would avoid what Sullivan has called "ontological expansion" (Sullivan, 2006) (the self-attribution by privileged agents of the right to participate in communicative scenarios that are modified by that participation), or what Bernstein has called "epistemic exploitation" (Berenstain, 2016), a phenomenon that occurs when the privileged alleged allies place the burden of proof and the responsibility to educate them on the shoulders of the discriminated and, in doing so, increase rather than decrease the epistemic gap.

Thus, if for veritism the reliability of a socio-epistemic practice can be measured by the ratio of true beliefs acquired to the total number of beliefs produced by that practice (Badhwar, 2009; Goldman, 2010), then lurking can be considered a reliable practice. The mere possibility of lurking would enable epistemic virtues such as open-mindedness, curiosity or humility. To my mind, there are two major objections to this view. (a) In my view, Frost-Arnold's analysis does not sufficiently explain whether these are virtues enabled by lurking, or whether they are character traits necessary for lurking to be virtuous. If the latter, as I am inclined to think, the argument would nevertheless rest on a responsibilist voluntarism. (b) Although the struggle against epistemic injustice associated with an action such as lurking involves a set of distinct virtues, in my view epistemic humility is the meta-virtue shared by all of them. Humility, for many, is more of a meta-virtue, because it is a willingness to revise our epistemic beliefs and attitudes when new evidence or testimony presents itself (J. S. Baehr & Hazlett, 2016). It thus entails a motivation to recalibrate our capacities, skills and experiences presupposed in inquisitiveness, curiosity or open-mindedness (the willingness to transcend the default standpoint and to take into consideration the merits of other standpoints) (J. Baehr, 2011, p. 152). According to Frost-Arnold's optimistic approach, epistemic life on the internet does not necessarily foster hubris but, by enabling specific dispositions such as those of the lurker, favours an intellectual humility that is not possible in offline environments.

In a very interesting twist of argument, Levy has argued that behind humility lies a presupposed arrogance: that of an epistemological individualism that assumes that we are always capable of autonomously revising our beliefs and that, ultimately, understands that only beliefs that have been subjected to critical examination qualify as knowledge (Levy, 2023). Whether for evolutionary reasons or because of enlightened cultural ascendancy, this

epistemological individualism has become the normative common sense of our everyday epistemic practices. For Levy, this contradicts the interdependent nature of our epistemic agency, which allows him to paradoxically interpret experiments that have provided empirical evidence about the alleged epistemic hubris behind the "illusions of understanding" (de Ridder). These experiments induce intellectual humility by showing participants that they were actually ignorant of how mechanisms they thought they knew work, or of policy measures they had chosen to support. The overarching effect is a distrust of prior beliefs and the emergence in participants of a more humble disposition towards their own cognitive capacities and epistemic resources. Levy does not evaluate these findings positively. On the contrary, in his opinion this demonstrates an epistemological individualism that pushes to revise unjustifiably (because of lacking the competences to do so) beliefs formed through justified deference. The paradoxical conclusion of this reasoning is that humility, instead of favouring deference, may on the contrary neutralise previous acts of virtuous deference.

This reasoning does not apply well to situations of structural epistemic injustice in which, precisely, the appeal to humility implies revisiting a deference that has had discriminatory effects. Nevertheless, I do find the critique of the underlying epistemic individualism useful: regardless of whether the prior deference was virtuous (Levy discusses situations in which humility can lead to refuting confidence in scientific consensus) or flawed (because beliefs loaded with discriminatory biases are accepted), the conclusion to be retained is that confidence in the possibility of an act of individual contrition is illusory. In other words, the online lurking of testimonies of discriminated people is not a sufficient guarantee for an individual to revise his or her previous assumptions, reformulate his or her beliefs and identify his or her biases. In fact, taking Levy's reasoning to the extreme, it could be irrational because it would push one to take excessive epistemic risks. This is why, in my view, trust and deference in testimony requires self-confidence that can only be provided relationally.

6. Coda: towards a model of group virtue?

My conclusion is that online lurking, although it may be a paradigm of how the Internet can trigger virtuous epistemic dispositions, does not have sufficient normative force to overcome the risks for the agent of trusting the testimony of subjects who, precisely because they suffer epistemic injustices, do not satisfy the requirements of trustworthiness. I believe that this conclusion authorises the leap to group virtues in search of a way to minimise the risks of this trust. This has been attempted by Lavinia Marin and Samantha Copeland, who have argued that self-confidence can only be born in communities that foster critical relational behaviour (Marin & Copeland, 2022). I believe that Mark Alfano's conception of communities of trust (Alfano, 2016), whose members share the meta-knowledge of reciprocal trust, is also useful for thinking about this leap to the group level. Although Copeland and Marin understand relational trust as a practice and Alfano rather as a shared certainty about the expected behaviour of others, both approaches agree in arguing that for it to be rational to trust an out-group witness, in-group trust has to be warranted. More specifically, Copeland and Marin describe trust in those with critical attitudes. The idea would be as follows: "All members of group G trust that if an agent X has a reasonable critical attitude about the group's position on issue A, and if he/she decides

to trust the testimony B of an agent Y external to the group, the members of group G must a priori trust X's good intention and X must know with certainty that the other members trust his/her action". This does not automatically imply transitivity, in the sense that because of that trust, the members of G automatically come to believe Y's testimony. This thesis would jeopardise the existence of the group. What this model seeks to guarantee is that the action of trusting and trying to change one's own belief system does not have an excessive cost, neither from the point of view of persuasive argumentation theory nor from the point of view of the motivations linked to group membership.

Although the purpose of this paper is not to develop that model, I believe that for the online lurking advocated by Frost-Arnold to be effective from a veritist (to produce and spread more true beliefs) and ethical (to take responsibility for one's own responsibility for certain injustices and commit to change oneself to mitigate them) points of view, group conditions must be in play to justify the rationality of these epistemic and ethical decisions. Otherwise, online lurking may be a way to alleviate the bad conscience of privileged actors at the cost of not triggering any commitment to transform themselves

References

Alfano, M. (2016). The Topology of Communities of Trust. *Russian Sociological Review*, *15*(4), 30-56. https://doi.org/10.17323/1728-192X-2016-4-30-56

Anderson, E. (2012). Epistemic Justice as a Virtue of Social Institutions. *Social Epistemology*, *26*(2), 163-173. https://doi.org/10.1080/02691728.2011.652211

Badhwar, N. K. (2009). The Milgram Experiments, Learned Helplessness, and Character Traits. *The Journal of Ethics*, *13*(2-3), 257-289. https://doi.org/10.1007/s10892-009-9052-4

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford university press.

Baehr, J. S., & Hazlett, A. (Eds.). (2016). The Civic Virtues of Skepticism, Intellectual Humility, and Intellectual Criticism. En *Intellectual virtues and education: Essays in applied virtue epistemology* (pp. 71-92). Routledge, Taylor & Francis Group.

Berenstain, N. (2016). Epistemic Exploitation. *Ergo, an Open Access Journal of Philosophy*, *3*(20201214). https://doi.org/10.3998/ergo.12405314.0003.022

Brinkmann, M. (2022). In Defence of Non-Ideal Political Deference. *Episteme*, *19*(2), 264-285. https://doi.org/10.1017/epi.2020.26

De Ridder, J. (2022). Online Illusions of Understanding. *Social Epistemology*, 1-16. https://doi.org/10.1080/02691728.2022.2151331

Driver, J. (2001). *Uneasy Virtue* (1.ª ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511498770

Dutilh Novaes, C. (2023). The (higher-order) evidential significance of attention and trust—Comments on Levy's *Bad Beliefs*. *Philosophical Psychology*, *36*(4), 792-807. https://doi.org/10.1080/09515089.2023.2174845

Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, *144*(3), 674-687. https://doi.org/10.1037/xge0000070

Fricker, E. (1995). Telling and Trusting: Reductuonism and Anti-Reductionism in the Epistemology of Testimony. *Mind*, *104*(414), 393-411.

Fricker, M. (2010). Replies to Alcoff, Goldberg, and Hookway on *Epistemic Injustice*. *Episteme*, *7*(2), 164-178. https://doi.org/10.3366/epi.2010.0006

Frost-Arnold, K. (2023). *Who should we be online? A social epistemology for the Internet*. Oxford University Press.

Goldman, A. I. (2010). Systems-oriented social epistemology. En *Tamar Szabó Gendler & John Hawthorne (eds.). Oxford Studies in Epistemology, Vol. 3.* (pp. 189-214.). Oxford University Press.

Heersmink, R. (2018). A virtue epistemology of the Internet: Search engines, intellectual virtues and education. *Social Epistemology*, *32*(1), 1-12.

Lackey, J. (2007). Why we don't deserve credit for everything we know. *Synthese*, *158*(3), 345-361. https://doi.org/10.1007/s11229-006-9044-x

Levy, N. (2019). Due deference to denialism: Explaining ordinary people's rejection of established scientific findings. *Synthese*, *196*(1), 313-327. https://doi.org/10.1007/s11229-017-1477-x

Levy, N. (2022). *Bad beliefs: Why they happen to good people* (First edition). Oxford University Press.

Levy, N. (2023). Too humble for words. *Philosophical Studies*, *180*(10-11), 3141-3160. https://doi.org/10.1007/s11098-023-02031-4

Marin, L., & Copeland, S. M. (2022). Self-Trust and Critical Thinking Online: A Relational Account. *Social Epistemology*, 1-13. https://doi.org/10.1080/02691728.2022.2151330

Sosa, E. (1995). *Knowledge in perspective selected essays in epistemology*. Cambridge University Press.

Sullivan, Emily. (2019). Beyond Testimony: When Online Information Sharing is not Testifying. *Social Epistemology Review and Reply Collective*, *8*(10), 20-24.

Sullivan, S. (2006). *Revealing whiteness: The unconscious habits of racial privilege*. Indiana University Press.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University press.

Véliz, C. (2019). Online Masquerade: Redesigning the Internet for Free Speech Through the Use of Pseudonyms. *Journal of Applied Philosophy*, *36*(4), 643-658. https://doi.org/10.1111/japp.12342

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 179-188
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.619861

*Epistemic communities and trust in digital contexts**

Comunidades epistémicas y confianza en contextos digitales

*ANTONIO GAITÁN TORRES***

Resumen: Este comentario se centra en la noción de 'comunidad epistémica' y en su rol para sustentar el argumento general que Karen Frost-Arnold presenta en *Who Should You Be Online?* (OUP, 2023). En la primera sección se presenta el argumento general de WSYBO, esbozando la estructura general del libro y sus conceptos centrales. En la segunda sección se distinguen tres sentidos posibles de 'comunidad epistémica' – sistémico, agregado y grupal. En la tercera sección se exploran las tensiones en torno al sentido grupal de comunidad epistémica. Se atenderá a dos formas en las que ciertas comunidades epistémicas cerradas y organizadas en torno a una identidad compartida pueden desviarse de los ideales epistémicos que guían el proyecto de Frost-Arnold. Algunas comunidades cerradas pueden organizarse en torno a dinámicas excluyentes. Otras comunidades pueden organizarse en torno a debates o controversias, afectando a la

Abstract: This commentary focuses on the notion of 'epistemic community' and its role in underpinning the general argument that Karen Frost-Arnold presents in *Who Should You Be Online?* (OUP, 2023). The first section presents the general argument of WSYBO, outlining the general structure of the book and its central concepts. In the second section, three possible senses of 'epistemic community' are distinguished – systemic, aggregate and group-oriented. The third section explores tensions around a variety of group-oriented epistemic community. It will address two ways in which certain closed epistemic communities organized around a shared identity can deviate from the epistemic ideals that guide Frost-Arnold's project. Some enclosed epistemic communities can be organized around exclusionary dynamics. Other enclosed epistemic communities may organize and grant membership around debates or controversies, affecting

Recibido: 25/04/2024. Aceptado: 25/06/2024.

* This paper has received the generous support of the Spanish Ministry of Science and Innovation Grants (METAPRODES – 'Meta-aptitudes, desacuerdos profundos y progreso moral' - PID2021-124152NB-I00) and the Fundación BBVA Research Scientific Projects 2022-24 (digi_MORALS - 'Los desacuerdos morales en la esfera digital – dinámicas interactivas, micro-mecanismos y marcadores culturales').

** Universidad Carlos III de Madrid, Profesor Titular, Departamento de Humanidades – Filosofía, Lengua y Literatura, Facultad de Humanidades, Comunicación y Documentación. Sus líneas de investigación incluyen la ética normativa y la meta-ética, la filosofía experimental y la epistemología social y política. Últimas publicaciones: Gaitán, A. (2024). "La psicología de las emociones políticas', en Gómez-Ramos, A. Velasco, G. (eds.). *Atlas de Emociones Políticas,* Madrid Trotta; Gaitán, A. Viciana, H. Aguiar, F. (2023). "The Experimental Turn in Moral and Political Philosophy", en Viciana, H. Gaitán, A. Aguiar, F. (eds.). *Experiments in Moral and Political Philosophy*, Londres, Routledge. Correo electrónico: agaitan@hum.uc3m.es

calidad deliberativa de esos debates. El potencial epistémico de las comunidades epistémicas cerradas y organizadas en torno a una identidad compartida depende en gran medida de evitar estas dos desviaciones.
Palabras clave: Confianza, comunidades epistémicas, epistemología social, epistemología orientada hacia sistemas.

the deliberative quality of those debates. The epistemic potential of closed epistemic communities organized around a shared identity depends largely on avoiding these two deviations.
Keywords: Trust, epistemic communities, social epistemology, system-oriented epistemology.

In *Who Should You Be Online* (OUP, 2023 – WSYBO hereafter), Karen Frost-Arnold proposes 'a socially situated epistemology for the internet' (WSYBO, p. 203); 'a philosophical, activist and non-ideal framework' (WSYBO, p. 5) that should contribute to improve our interactions in various digital contexts. To articulate this framework, Frost-Arnold appeals to several philosophical concepts, as well as a huge wealth of empirical evidence focused on our behavior and attitudes in digital contexts. WSYBO is full of concrete examples, always illustrating far-reaching philosophical claims in an entertaining and accessible way. All this is achieved in a constant dialogue with disciplines such as Anthropology, Sociology, Political Science, Communication or Digital Studies.

There are many interesting aspects of WSYBO that could be commented on, either on some long-standing debates in Epistemology (trust, epistemic autonomy, epistemic injustice), or on more novel topics and lines of research (the epistemic risks of moderation, the group biased nature of fake-news, the dangers and possibilities of imposture in digital contexts, 'lurking' as an epistemic practice and a long etcetera). Here I will focus on the notion of 'epistemic community', a concept that I find especially interesting when it comes to understanding and evaluating the project that Frost-Arnold develops in WSYBO.

1. The project

The central thesis of WSYBO is that any 'epistemic system' (press, education, social media, etc.) that seeks to promote the generation of objective knowledge – free of biases and beyond partial perspectives – must be capable of implementing two great ideals. First, it must be diverse in a strong sense, that is, the epistemic system must prioritize not only a plurality of perspectives; It must also include the perspective of the disadvantaged or oppressed groups. Secondly, and as a necessary condition for getting the epistemic benefits to be brought by this strong diversity, it must promote and protect networks of trust that ensure the expression of subaltern or oppressed voices (WSYBO. p. 75).

It is convenient to stop at this second condition. According to Frost-Arnold, diversity in the strong sense has no real effects for the generation of objective knowledge if oppressed groups are not part of the trust networks established between different epistemic communities. For the inclusion of oppressed voices to be effective, trust networks must be robust in three different senses. Oppressed groups must trust their own perspective, have the trust of other members of the community, and themselves trust those practices that the community establishes to encourage the debate and the inclusion of different voices and perspectives (WSYBO. p. 76).

Assuming this general framework, much of WSYBO can be read as a catalog of characters or ideal types that would undermine networks of trust at one of its three vertices. Chapter two, for instance, explores a variety of 'imposture' in digital contexts in which an oppressed perspective is supplanted with the goal of fostering stereotypes about such perspective. During the Arab uprisings of 2010, a blog attracted the attention of many Western journalists and political commentators. The title was very suggestive: *A gay girl in Damascus* told the story of the Syrian revolution from the perspective of Amina Arraf, a lesbian activists of Syrian-American origin. Amina narrated in first person her experiences during the uprisings, giving a first-hand account of the ongoing political insurrection. Months later it would be discovered that the blog was actually a fake and that its author, Tom MacMaster, was an American student living in Scotland. MacMaster's imposture, regardless of his real motivations, surely eroded the confidence of many of the readers of the numerous political activists blogs that reported on the ground and in real time about what was happening in Syria. And this impoverished the understanding of the Syrian political uprising during those years.

WSYBO caters in several characters like this, always asking about their effects on those networks of trust necessary for the production and dissemination of objective knowledge. Attention is also paid to the role that the architecture and affordances of digital environments would play in protecting or eroding these networks of trust. The possibility of interacting in spaces protected by anonymity, or the possibility of controlling who participates in those forums in which oppressed groups are discussing problems related their oppression, allows such minority groups to safely develop critical and emancipatory perspectives. The last chapter, dedicated to lurkers, offers an example of these contexts and dynamics.

Frost-Arnold's project is carried out by taking for granted several concepts and debates. Within this territory of implicit assumptions, there is a word that appears a lot of times in WSYBO. I am referring to the notion of 'epistemic community'. Almost all the actors who appear in this book are not interacting in the vacuum. By contrary, they are inserted, more or less formally, into different epistemic communities. But what is an epistemic community? And how are epistemic communities fitting into the general project outlined above? Do the trust networks that enable the generation of objective knowledge depend on a certain variety of epistemic community? And if so, how to understand that dependency? I am sure the author can answer all of these questions, but even so let me develop here some tentative criticisms for the sake of having a better picture on the meaning and varieties of epistemic communities in WSYBO.

2. Three senses of epistemic communities

Three broad senses of 'epistemic community' can be traced throughout WSYBO. According to the first, *a systemic or structural sense*, an epistemic community is equivalent to a socio-epistemic system of rules, conventions, techniques, affordances and practices that are organized (at least partially) for the promotion of knowledge. Facebook and X are epistemic communities in this structural sense. They stand for digital spaces

designed and organized, at least in part, with the aim of sharing information and promoting debate - two goals that are essential for promoting objective knowledge in Frost-Arnold' sense[1].

In a second sense, epistemic communities are sometimes referred in WSYBO as *aggregates of opinions around a particular debate or topic*. Some threads on X in which disagreements develop over some public controversy offer a clear example of an epistemic community in this second sense. In these aggregates of opinions, sometimes very numerous, interaction is 'more or less' limited to the exchange of points of view on a given topic, for example, police brutality (WSYBO. p. 60). These epistemic communities can persist over time, forming stable spaces in which we can find and supply information on a certain topic.

Finally, in WSYBO a third sense can be traced, according to which an epistemic community refers to *a group organized around a topic, debate, controversy or shared social identity*. This is the sense that I am interested in exploring in this commentary. Now we are not referring to a mere aggregate of opinions, as in the previous case, but rather to spaces or communicative structures in which a *group* of people interact on a regular basis and in which we find different degrees of organization. The organization of these groups can be informal or regulated, allowing for different levels of openness and inclusion (WSYBO. p. 161). The stability of the interactions between members of an epistemic community can also vary, moving form longstanding groups to more momentaneous assemblies. And importantly, epistemic communities in this 'groupish' sense can be organized around a topic or interest, but also around a shared social identity.

A blog about the Second World War in which a significant number of users participate regularly constitutes an epistemic community in this third sense, that is, a group organized around a specific *topic*. The limits between the second sense seen above and the one we are now considering are sometimes blurry, but certain features can give us a clue as to when we are dealing with an epistemic community organized as a group and when we are facing an aggregate of opinions around an issue. The administrator of the blog we are considering could, for example, control who participates or instead have a more open policy, allowing to anyone who is interested in the topic to post opinions in the blog. To the extent that the organization of a blog allows the administrator to control who post on the topic, a blog is clearly different from a thread of X on the same topic. Frost-Arnold writes:

> "(…) blogs can create smaller epistemic communities by hosting active comment threads. Commentators often engage in a critical dialogue with one another, another potential objectivity enhancing practice". (WSYBO. p. 80)

In other epistemic communities, by contrary, the criteria for belonging do not have so much to do with specific debates or topics as with *explicit affiliation to a shared identity*. A group of black women who start a blog to share their experiences of discrimination in

1 The interesting discussion around moderation contained in the second chapter refers to this structural sense of epistemic community (p. 32). The epistemic risks of moderation (the silencing of minority voices, for example) also naturally refer to this structural or systemic level. The numerous references to inequalities between groups or the underrepresentation of a certain community within a social network are also examples of this first sense of epistemic community (p. 34)

the workplace and who are reluctant to allow the participation of non-black women offers a clear example of this type of epistemic communities (WSYBO. p. 170).

Communities organized around a shared social identity appear in various parts of WSYBO. The assumption is that, at least when they are not organized in an excessively closed or opaque way (allowing some non-black women to participate, or facilitating the 'lurking' of their interactions by external observers), they enact spaces that would facilitate the generation of objective knowledge (WSYBO. p. 166). In the case we are currently focusing on, the epistemic community of black women would offer first-hand information about the living conditions and experiences of an oppressed group. Lurkers could access this information and transmit it to other epistemic communities, enriching the overall debate (in the systemic sense noted above) with information about the group's effective situation of exclusion, or about their own perception of that situation. This information may be essential to guide and modulate possible policies aimed at mitigating or correcting such situation. Even if interactions with other epistemic communities are somehow reduced, these closed epistemic communities organized around identity-markers can generate long-term epistemic benefits. By offering security to its members and a basic feeling of understanding, they allow that certain experiences can be articulated more precisely.

These three varieties of epistemic communities appear in WSYBO. And the central argument outlined above can, in fact, be refined by considering these three broad varieties. The generation of objective knowledge evaluated in a systemic way points to the systemic sense outlined above. Diversity of opinions, in the strong sense that concerns Frost-Arnold, can be exemplified in specific contexts in which an issue is debated (second sense of epistemic communities), but it can also refer to the composition of groups with different degrees of opening (third sense of epistemic community). And the same can be said about trust, the oil that allows that diversity can have real effects on the promotion of objectivity. Sometimes, trust has to do with the ability to intervene in open deliberative spaces without suffering verbal attacks or threats (second sense). Other times, however, trust requires the protection offered by more enclosed groups, which allow certain critical opinions to mature and be articulated effectively (third sense). As I noted above, one of the most valuable things of WSYBO is the way it precises how these enclosed groups could also be of critical importance for generating objective knowledge in the broader epistemic community (first sense).

At a more general level, the evolution of our perception of digital contexts, from a distant period of initial optimism to the current wave of pessimism (which Frost-Arnold dates to around 2016), can also be interpreted in light of the three senses of epistemic community outlined above. The initial promises undoubtedly pointed to the potential of digital forums, understood as open spaces of deliberation, to generate objective and open knowledge (first and second sense) (Sunstein 2009). The current pessimism around social media points, by contrary, to the proliferation of powerful groups interfering with the free exchange of perspectives and ideas (third sense) (Habgood-Coote 2024).

As I have suggested above, Frost-Arnold assumes part of this general narrative about our perception of the potentialities of digital spaces, but she does not endorse an additional move, common in many conservative analyses of the epistemic potential of the internet. Frost-Arnold does not believe that closed groups organized around identity markers are necessarily positing a danger to the generation of objective knowledge. Under a systemic

perspective they can promote objective knowledge. This precise claim, framed in epistemic terms, is one of her major contributions in WSYBO.

In what follows I will focus on some epistemic communities organized around a shared identity. Although Frost-Arnold is correct when she accentuates the epistemic advantages of some of these communities, her discussion blurs some risks associated with other possible instances of identity-based epistemic communities. I will argue that keeping these risks in sight could help us to better understand the potential of Frost-Arnold's general argument. The goal of the following section, in any case, is mostly exploratory. I am sure the author can deal with most of the criticisms I'll unfold.

3. Identity-based epistemic communities - the good, the bad, and the ugly

There is a general and straightforward way to articulate what concerns me in relation to the 'groupish' and identity-related sense of epistemic community. This strategy would point out, in a nutshell, that our group tendencies involve dispositions and beliefs that have bad epistemic consequences. Favoring the members of our group, limiting contact with members of other groups, openly discriminating against them when distributing resources, time and attention, are paradigmatic examples of epistemically deviant 'coalitional' dispositions (Boyer 2020). Although there are some senses in which some group-centered bias can have epistemic value (Rini 2017), coalitional dynamics have in principle little value in generating the kind of objective knowledge that is put at the center of WSYBO. Therefore, any notion of epistemic community that is articulated around group-based categories must explain how it would redirect the dynamics of exclusion and isolation typical of groups in order to ensure that the systemic epistemic benefits end up being positive. Are there any features of the epistemic communities exemplified in WSYBO that can minimize the corrosive nature of group dynamics? How to prevent the exclusion of perspectives? How to limit the polarization of opinions? How to minimize tendencies such as prejudices and biases that seem to prevent the articulation of objective knowledge? Does Frost-Arnold have in mind some cultural factors, organizational elements, rules or social norms that would minimize the effect of group dynamics on identity-related epistemic communities?

As I said before, the above offers an initial and sketchy route to question the epistemic potential that Frost-Arnold assigns to isolated epistemic communities organized around a shared identity. In what follows, I will try a second critical strategy, which involves exploring other possible varieties of identity-based epistemic communities. I think that keeping these possibilities in mind – and some of their negative outcomes and traits – can help us relativize Frost-Arnold's optimism in relation to epistemic potential of identity-based epistemic communities.

Let's go back to Frost-Arnold's example of a closed epistemic community organized around an identity marker:

> **(EC $_{good}$)** A blog for black women organized to share their experiences of discrimination in a specific work context. The group does not allow non-black women to participate (although it does allow 'lurking').

Closure to other perspectives helps members of this community to better articulate a particular perspective of oppression. Let's compare this community with two other varieties of epistemic communities organized along identity lines. Here is an example of what we could call an 'ugly' epistemic community:

> **(EC $_{ugly}$)** A blog in which a group of mothers interact regularly about the risks of vaccines. The group allows any user to participate, although most interactions are structured around the values and beliefs shared by a stable group of mothers exchanging views on the topic in question.

And here is an example of a harmful or 'bad' epistemic community:

> **(EC $_{bad}$)** A community of X users organized with the aim of marking as discriminatory or harmful any protest or critical content related to the interests of an oppressed group (p. 34).

Why does a group of black women who share their experiences of discrimination constitute a 'good' epistemic community, while a group of mothers who exchange views on the risks of vaccines offer an example of an 'ugly' epistemic community? And what separates these communities from those that are clearly harmful or 'bad'? Without aiming to be systematic (and with the intention of knowing Frost-Arnold's take on these cases) I think that several interesting things can be said about these cases. Some of these things could help us to understand the possible (and frequent) deviations that beset the 'good' epistemic communities highlighted by Frost-Arnold in WSYBO.

Let's start with the easy case. There is a clear sense according to which **(EC $_{bad}$)** is a 'bad' epistemic community. It actively discriminates against members of other groups, favoring negative attitudes and emotions towards those people. These negative attitudes undoubtedly hinder access to the epistemic goods derived from contrasting opinions through open debate and inclusion. It could be claimed that in these cases the shared identity markers are framed 'negatively', that is, in opposition to other groups towards which discriminatory attitudes are actively promoted. These communities would be close to what Nguyen has labelled as 'echo chambers'. An echo chamber is a communicative structure in which other perspectives are let out *and actively discredited* (Nguyen 2018). When we keep this case in mind, it sounds reasonable to claim that, as least as a general rule, good epistemic communities must avoid fostering negative attitudes toward other groups. These negative attitudes could prevent the group from getting the epistemic benefits derived from interaction and openness. Good epistemic communities organized around a shared social identity would embrace identity positively, that is, as a criterion of belonging but never in opposition to another group.

Let's consider now the 'ugly' case. This one is a little more interesting. In (EC $_{ugly}$) the group of mothers structures their interactions around a topic, the risks of vaccines, which constitutes the focus of a more general debate, present in other forums and contexts and largely governed by experts. An 'ugly' epistemic community does not actively exclude other groups, and in this particular sense it is clearly different from the 'bad' communities sketched above. However, this type of communities differs from (EC $_{good}$) in a different sense, one that

seems very relevant and that would point towards another possible deviation to be avoided by a good identity-based epistemic community.

In Frost-Arnold's favored example, a group of black women organize their interactions with the goal of voicing or expressing their experiences in a certain work context. What characterizes these closed informational structures has to do in an important sense with the *security* they offer to their members to share these experiences - experiences endowed with epistemic value because they are peculiar and not mediated by relations of oppression, threats, etc. (Furman 2022). Keeping that general point in mind, the second deviation that I want to point out occurs between epistemic communities in which the identity marker is made explicit and puts us on the track of a forum in which we are going to learn things about the experience or experiences. of a certain group (EC_{good}) and those other communities focused on a particular topic or controversy (vaccines, climate change) and organized through non explicit dynamics of belonging and group affinity (EC_{ugly}). The risk with this second type of epistemic community is that now the usual groupish dynamics that we can find in any group can affect to how the group, as a deliberative system, would approach to the evidence and arguments surrounding the topic in question.

Dynamics of polarization and extremism, confirmation biases from dogmatically defended positions, policing of deviant opinions, and other similar phenomena could potentially affect the quality of the deliberation of these groups. Now it is no longer about expressing a perspective of oppression or a peculiar experience, the epistemic center of 'good' epistemic communities. Now it is about intervening in more general debates from a set of groupish dynamics inserted in closed informational structures.

Dan Williams has recently described these epistemic communities as groups in which a specific topic becomes a flag of group membership (see also Van Leeuwen 2023). Williams suggests that such groups, organized around certain 'identity-defining beliefs', exemplify processes of group epistemic deliberation of lower epistemic quality and negative aggregated effects. Williams mentions at least three of these processes: selection of evidence in line with the beliefs that define the group, dogmatic defense of those beliefs and what he refers to as 'markets of rationalizations' (Williams ms. See also Mercier & Sperber 2017. Gaitán 2024).

Above I asked if we can say something positive about the features that would separate 'good' epistemic communities organized around a shared identity from harmful or 'bad' epistemic communities (also sharing a common identity) and from those that I've referred as 'not so good' or 'ugly' epistemic communities. Now we can at least specify two general traits to be avoided:

— A good epistemic community organized around a shared identity must avoid defining itself negatively, that is, in opposition to other identities which are denigrated, insulted or stereotyped.
— A good epistemic community organized around a shared identity must protect a space in which experiences are expressed safely, avoiding taking a topic or debate as a flag of membership.

Bad epistemic communities organized around a shared identity are not necessarily bad because they limit interaction with alternative views. One virtue of WSYBO is precisely to remove this assumption, making space for a more positive view on the epistemic impact of a certain variety of closed epistemic communities. But epistemic communities organized

around a shared identity can be deviant for additional reasons beyond closeness. Here I've tried to highlight two of these reasons. They can be exclusionary and they can promote deliberative spaces where some topics or debates are behaving as flags of membership. There may be more negative features of identity-based epistemic communities, but I believe that the two I've just highlighted are especially salient in current digital contexts.

A precautionary note is in place. Maybe the image I am sketching here is too narrow to catch the political potential of closed (and good) epistemic communities. After all, there is a sense in which I am claiming that if they want to be epistemically fruitful, these communities must limit themselves to the articulations of experiences of oppression, avoiding wider political issues or debates, usually polarized and extremely divisive. This can be interpreted as conservative, even as regressive. But there are some tools in WSYBO that could help us to minimize this problem. We could secure the political potential of good epistemic communities by promoting lurking, or by establishing some deliberative satellites that could serve as a bridge between closed and good epistemic communities and the wider deliberative and political space (Squires 2002).

Conclusion

One of the virtues of WSYBO is the careful epistemic defense of closed communities organized around identity markers that it offers. In this commentary I have tried to specify the epistemic value of these communities, which according to Frost-Arnold has to do with the articulation of secure spaces in which oppressed groups are able to express their experiences. These shared experiences can influence the epistemic quality of the wider debate through various indirect means, from direct lurking to the articulation of regulated interactions with other communities or forums. However, the groupish nature of these communities also implies that their 'deliberative' focus must be kept very precise: it is advisable to avoid their interactions being articulated around specific themes or debates and in no case should they be defined negatively, as in explicit opposition to other groups. It is in the narrow margin where experiences of oppression are expressed where the value of these closed epistemic communities must be located.

References

Boyer, P. (2020). *Minds Make Societies. How Cognitions Explain the World Humans Create*, Yale University Press.

Furman, K. (2022). 'Epistemic Bunkers', *Social Epistemology*, DOI: 10.1080/02691728.2022.2122756

Gaitán, A. (2024). 'La psicología de las emociones políticas', en Gómez-Ramos, A. Velasco, G. (eds.). *Atlas de emociones políticas*, Madrid, Trotta

Habgood-Coote, J. (2024). 'Toward a Critical Social Epistemology of Social Media', in Lackey, J. McGlynn, A. (eds.). *Oxford Handbook of Social Epistemology*, Oxford University Press.

Mercier, H. Sperber, D. (2017). *The Enigma of Reason*, London, Penguin

Nguyen, C. T. (2018). 'Echo Chambers and Epistemic Bubbles', *Episteme*, doi:10.1017/epi.2018.32

Rini, R. (2017). 'Fake News and Partisan Epistemology', *Kennedy Institute of Ethics Journal*, 27 (52), pp. 43-64

Squires, C.R. (2002), 'Rethinking the Black Public Sphere: An Alternative Vocabulary for Multiple Public Spheres'. Communication Theory, 12: 446-468. https://doi.org/10.1111/j.1468-2885.2002.tb00278.x

Sunstein, C. (2009). *Republic 2.0*, Princeton University Press.

Van Leeuwen, N. (2023). *Religion as Make Believe. A Theory of belief, imagination and group identity*, Harvard University Press.

Williams, D. (ms). 'Identity-defining beliefs'

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 189-196
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.620301

Response to Comments

Respuesta a comentarios

KAREN FROST-ARNOLD[*]

I am grateful to the authors in this symposium for such thoughtful and thought-provoking analyses of *Who Should We Be Online?* Each piece has helped me better situate the book and clarify my own ideas about what I tried to achieve in this project. Beatriz Jordá's piece usefully frames the book as wrestling with optimism and pessimism about the internet as a tool to combat ignorance. On the one hand, the internet is a space of frenetic epistemic activity—so much of what we learn today is discovered via the internet. And, as Jordá points out, the internet can connect us to people very different than us. This provides opportunities for learning about social justice, unlearning our own biases and stereotypes, and developing virtuous epistemic habits of listening to people who experience oppression. Thus, there are many epistemic benefits from our online lives. On the other hand, the internet is rife with disinformation, hoaxes, and bad actors. I appreciate that Jordá recognizes both my theoretical and practical goals in the book. Theoretically, the book develops a socially situated epistemology with a set of interdisciplinary tools for analyzing the promises and perils of the internet. My goal is not to address every question about internet epistemology, but rather to demonstrate the value of these particular theoretical tools by applying them to several online personas (moderators, imposters, tricksters, fakers, and lurkers). As Jordá notes, the book also tries to help each of us navigate the challenging online landscape. We can think about how our online activities promote objectivity, truth, and epistemic justice, and we can try to cultivate habits for unlearning our biases and ignorance. Since the publication of *Who Should I Be Online?*, I have been particularly excited to see developing work by sev-

[*] Professor of Philosophy, Hobart & William Smith Colleges and Visiting Associate Professor, the African Centre for Epistemology and Philosophy of Science, University of Johannesburg. Her main lines of research are ethics, philosophy of science, feminist epistemology and social epistemology. Recent publications: Frost-Arnold, K. (2021). "The Epistemic Dangers of Context Collapse Online." In *Applied Epistemology*, (ed.) J. Lackey. New York: Oxford University Press; Frost-Arnold, K. (2020). "Trust and Epistemic Responsibility." In *The Routledge Handbook of Trust*, (ed.) J. Simon. New York: Routledge.
frost-arnold@hws.edu

eral philosophers using the research ethics appendix of the book to work through the tricky problems of how to conduct research about the internet in an ethical way.

Lola Medina Vizuete's piece makes an excellent contribution to the literature on online ignorance by noting a gap in my analysis of miscognition in our collective social memory. In the book, I drew on Charles Mills' concept of white ignorance to show how fake news caused by white racism and white racial domination shapes our understanding of our past (Mills 2007). In "White Ignorance," Mills explains white ignorance as *false belief* or *absence of true belief* caused by white racism. I argued that when search engine archives store racist fake news, they help maintain racist false beliefs, and this is a form of digital white ignorance. Thus, the book explores how the dissemination and storage of false claims online causes ignorance. However, as Medina Vizuete notes, I did not investigate the other type of ignorance suggested by Mills: the absence of true belief. I did not examine whether white racism prevents certain types of knowledge from circulating online or remaining in online archives. Medina Vizuete's piece brilliantly presents several ways that the absence of certain voices online might cause gaps in our collective social memory. Medina Vizuete is certainly right that minoritarian beliefs are less likely to be shared online, and I think she persuasively argues that non-propositional knowledge is also under-represented online. I will present some examples that I think further support and extend her argument.

First, Medina Vizuete argues that small minority groups may have their beliefs proportionately represented. Due to their small number, minoritarian claims may be swamped out by the majoritarian view of the collective social memory. This strikes me as an important mechanism for the creation of gaps in our collective knowledge, and significantly Medina Vizuete points out that this epistemic problem falls outside the scope of epistemic injustice—insofar as the minoritarian groups are not being unjustly discriminated against as knowers. As Medina Vizuete points out, standpoint theory shows that minority groups often have valuable insights that should be attended to by the population at large. It is important that we create online spaces in which these kinds of minoritarian knowledges can be created and heard. For me, the challenge of finding ways for the insights of minority populations to be collectively generated, disseminated, and given uptake motivates the argument for online spaces where these populations can gather to create knowledge and plan for its dissemination. This is a topic which Medina Vizuete and Barbarrusa have skillfully addressed in a paper where they argue for the value of epistemic bubbles to foster the knowledge of patients with rare diseases, such as Cystic Fibrosis (Medina Vizuete and Barbarrusa Forthcoming).

Second, Medina Vizuete points out that lack of resources, education, and access to the internet may prevent some groups from adding their beliefs to the collective social memory. This is an important point. If applied epistemologists hope to provide recommendations to improve the quality of knowledge online, then we ought to consider the need for redistribution of wealth and resources. This is particularly important when we think about calls for epistemic decolonization (cf. Mitova 2020; Tobi 2020). To undo centuries of colonial domination of knowledge production and the suppression of Indigenous knowledge requires not only the removal of colonial influences, but also the "proactive utilisation of the marginalised epistemic resources of the colonised in the advancement of knowledge in various fields" (Mitova 2020, 192). In order for these epistemic resources to be used online, colonized communities need internet access, education, and other material

resources. Thus, global redistribution of wealth and resources has epistemic merit. I think this point is also relevant to Medina Vizuete's concerns about the absence of some types of non-propositional knowledge online. While Medina Vizuete is right that knowledge-how can be resistant to testimonial transmission, it can be learned though digital visual media, such as YouTube videos (Nagel 2017). However, if minority communities lack access to digital technologies to produce and post such videos, some of their knowledge-how may not be well represented online.

Third, Medina Vizuete draws our attention to groups who choose not to participate in online conversations and thus deprive the community of their knowledge. This deprivation can also skew our collective social memory. As Medina Vizuete notes, there are many reasons why people choose not to share their knowledge online. One reason that more social epistemologists should consider is privacy. Fear of online harassment, doxing, having one's information sold to data brokers, and dislike of surveillance by advertisers and others online are all pressing privacy considerations that push people to stay off social media (Citron 2022). When applied epistemologists consider social, legal, and political suggestions to increase the diversity of knowledge online, greater protections for privacy ought to be part of the conversation, in order to encourage a wider array of minoritized groups to share their knowledge online.

Gonzalo Velasco Arias' contribution raises important questions about epistemic virtue and individualism. Velasco Arias focuses on my use of situated virtue epistemology and systems-oriented social epistemology. Bringing these two approaches together asks us to identify how the social-epistemic structure of the internet shapes (and is shaped by) users' virtues and vices. Velasco Arias finds this approach useful for avoiding both doxastic voluntarism and a kind of naïve faith in individual responsibility. This is compatible with one of the goals of the book: to argue that the way to epistemically improve the internet is not simply to expect individuals to choose to do better and act more responsibly. Structural changes are necessary, and structural changes can shape who we become as online knowers. Nonetheless, Velasco Arias finds a lingering individualism in my account of epistemic virtue, and I find his analysis very helpful. His argument begins by drawing our attention to two features of online testimony that pose evaluative challenges for users. First, people can pretend to be someone they are not online (Frost-Arnold 2014; 2023). Second, the internet can give users an illusion of understanding that encourages vicious epistemic arrogance (Levy 2019; de Ridder 2022). Velasco Arias argues that these two problems make the internet a hostile epistemic environment for agents. Thus, we should be skeptical about my argument in chapter 5 that privileged people can gain knowledge from marginalized people's online testimony. I argued there that virtuous lurkers can responsibly gain knowledge by listening to marginalized people, but Velasco Arias is pessimistic about this possibility. He asks for clarification about my notion of the virtuous lurker. Are individual agents virtuous lurkers because they possess other virtuous character traits, or does the structure of the internet shape lurkers to develop the relevant virtues? He worries that if my account rests on the former claim, then my account devolves into a voluntaristic responsibilism. But if my account takes the latter route, then it renders epistemic agents irrational by assuming they take unjustified epistemic risks by believing marginalized agents online.

This is a very insightful argument, which helped me clarify my own thoughts about nature of epistemic virtue in the book. The book recognizes the importance of structural changes to improve epistemic virtue, but it also identifies a role for individual epistemic responsibility. Thus, the chapter on lurking argues that privileged people need to choose to be more careful in spaces where marginalized people gather. Velasco Aris may view these parts of the chapter as unacceptably responsibilist. I do not eschew all forms of responsibilism. And while I reject a naïve faith that individual responsibility can solve all our epistemic problems online, I do think agents can work to develop virtuous habits. Next, I want to focus on his comments on epistemic risk, pessimism, and deference.

First, let's begin with the idea that the internet is a hostile epistemic space. I do not think philosophers are in a position to make this kind of claim today. It is certainly true that we have many theoretical tools to identify epistemic challenges of the internet. Velasco Arias successfully presents two such problems (imposters and the illusion of understanding), and my book acknowledges others. However, the internet also provides many epistemic benefits. Thus, to know whether the internet is a hostile space, we need to know whether the benefits outweigh the costs. And to answer that question, we need to both specify what we take to be the measure of epistemic value (something philosophers can do) and also to have data about what is happening online (something we need social scientists and others to do). For example, suppose we agree with my account that truth is epistemically valuable and that internet imposters often spread falsehoods. Then, to know whether the internet is an epistemically hostile space, we need to know many imposters there are, how influential are they, how many false beliefs do they propagate, and how much do they undermine trust? I do not think we have adequate data to answer these questions. Additionally, I am not sure it is the best approach to ask questions about how hostile or risky the internet is as an overall epistemic environment. The internet is an incredibly diverse set of overlapping epistemic spaces. Wikipedia has grown to become a relatively reliable epistemic tool (Frost-Arnold 2018), but Truth Social is a toxic epistemic space. And individual platforms can change over time. For example, Elon Musk has removed many of the guardrails on Twitter (now X). At this point in time, I do not think philosophers have developed the collaborations with researchers in other fields (or have developed the abilities ourselves) to answer the empirical questions we would need to make assessments about whether a particular epistemic space at a particular point in time is hostile. My book aims to provide situated theoretical tools that would help us formulate questions and develop a better understanding of the kinds of collaborations necessary to answer them in the future. I think Velasco Arias' argument here relies on an empirical claim that believing marginalized people's testimony via lurking requires excessive epistemic risk, and I am not convinced that this is true.

Finally, I found Velasco Arias' comments on deference and the need for self-confidence to be relationally grounded very interesting. However, I am not sure that analyzing the kind of epistemic acts that lurkers engage in as deference is the most helpful framing. In the chapter on lurking, the epistemic agents who I argued would benefit most from lurking are agents already beginning to unlearn their socially situated ignorance. These are privileged people who realize that they have some privilege and that they may be unaware of its scope. They suspect that they may have some false beliefs as a result of their social location, but they do not know how many or what they are. For people at this point in their journey in

unlearning their ignorance, lurking in spaces for marginalized people may be useful, but I do not think they are simply deferring to the judgment of marginalized people, in Velasco Arias' sense. Velasco Arias adopts a broad definition of deference: "A defers to B on the question whether p if A believes that p (or not-p) merely because B believes that p (or not-p)" (Brinkmann 2022, 267). For him, deference is a radical act of trust in the speaker's trustworthiness. This is not what I think is going on at this stage in the process of unlearning one's ignorance. The privileged person does not simply defer to a marginalized testifier. This is less of a matter of replacing previous beliefs with new beliefs based on testimony, and more of a matter of raising questions and doubts and reasons to look for further evidence for one's beliefs. Part of what is going on in this process is helpfully illuminated by Karen Jones' account of a metastance of distrust in one's patterns of distrust (Jones 2002). Jones argues that in the process of unlearning their ignorance, epistemically responsible agents can cultivate habits of reflecting on their patterns of distrust (Jones 2002, 166). Suppose that I, as a white person, notice that I have a pattern of distrusting people of color, and this distrust is best explained by stereotypes and prejudices. Then I ought to adopt a metastance of distrust in my distrust. In other words, I ought to distrust my tendency to distrust people of color. Instead of arguing that I ought to just defer to their judgment about certain questions, Jones argues that this metastance of distrust will push me to look for more evidence. The more I distrust my own distrust, the less weight my judgment that the speaker is untrustworthy has and the more corroborating evidence I ought to seek of the agent's trustworthiness (Jones 2002, 164–65). Thus, by developing habits of reflecting on their prejudices, epistemically responsible agents can adopt attitudes towards their own trust that put them in epistemically better positions to judge when their rejection of testimony requires more evidence.

Lurking in marginalized communities provides opportunities to find such evidence. Lurking is not simply reading one tweet by a marginalized person and deferring to their judgment. Instead, it is a persistent practice of listening. And listening to many voices of a community with a metastance of distrust towards one's own potentially biased habits can put one in a position to gather more evidence that over time will lead to a more virtuous testimonial sensibility, in Fricker's sense (Fricker 2007). Again, how epistemically useful any particular community or space is should be determined case by case. The goal of the book was not to argue that believing the testimony of marginalized people online is always warranted, but instead to draw our attention (as both individual users making decisions about where to lurk online and also as philosophers evaluating the merits of different acts of lurking) to the relevant features of the online space, structures, and habits of members of the community.

Antonio Gaitán Torres' contribution helpfully examines the notion of an epistemic community. He asks for clarification about what distinguishes epistemically good identity-based epistemic communities from bad ones. He claims that "favoring members of our group, limiting contact with members of other groups, openly discriminating against them when distributing resources, time and attention" are "paradigmatic examples" of epistemically harmful group behavior (Gaitán Torres 2024). Noting that I argue for the epistemic benefits of online communities in which members of marginalized groups have space to talk to each other, he asks what distinguishes such groups from epistemically toxic communities, such as anti-vaxxer communities or groups targeting marginalized people for online harassment

(e.g., white users who report people of color discussing racism for being anti-white). These are useful questions, and I appreciate the opportunity to expand on these points.

First, it is crucial to recognize that much social epistemological work on epistemic groups has failed to take the socially situated approach that I pursue in the book. Mainstream social epistemology has taken the generic individual knower stripped of their social location as the primary entity of analysis, thereby failing to recognize that knowers have differing degrees of power and privilege. Similarly, when the field has turned its attention to groups as the primary entity of analysis, it has also failed to recognize the diversity of groups and the epistemic significance of how much power groups of knowers have. As decades of feminist epistemology and epistemology of race have shown, when this diversity is erased, the paradigmatic generic individual knower (or paradigmatic generic group of knowers) tends to resemble privileged individuals (or groups). For example, we end up with accounts of knowers that resemble white men, and accounts of groups of knowers that resemble groups of white men. For this reason, I think we should be very careful when we draw conclusions from what have been taken to be "paradigmatic examples" of epistemic risks of groups. These are likely to be risks of epistemically harmful behavior by privileged groups, but not necessarily epistemically harmful behavior by groups of oppressed people. In fact, a socially situated epistemology can recognize what is often obscured by mainstream epistemology—that behaviors such as favoring members of our group or limiting contact with members of other groups are actually epistemically beneficial for some groups but not others. Why is this? Two sets of insights from the epistemologies of ignorance and standpoint epistemology provide answers.

First, oppressed groups live in a world that systematically denies, undermines, and erases their knowledge. The epistemologies of ignorance literature has uncovered many mechanisms by which this occurs, including white ignorance, testimonial injustice, willful hermeneutical ignorance, gaslighting, testimonial smothering, to name just a few (Mills 2007; Fricker 2007; Dotson 2011; Pohlhaus Jr. 2012; McKinnon 2017). As my book argues, these ignorance-producing practices are often operative online. Additionally, marginalized people are disproportionately targeted for online harassment and abuse. This systemic epistemic oppression marks the first key difference between the "good" epistemic community Gaitán Torres imagines (the blog for Black women discussing their experiences with racism) and his "bad" and "ugly" communities (the anti-vaxxer mom blog and the group organizing to attack oppressed groups). The former group experiences pervasive attacks on its ability to produce and share knowledge, while the latter two do not. For this reason, it is epistemically beneficial for marginalized people to engage in behaviors that mainstream epistemology has improperly labelled as paradigmatically harmful. It is epistemically helpful to withdraw from people who systematically gaslight us, interrupt us when we speak to tell us that we are not competent, who share stereotypes that we are dishonest, or whose ignorance is simply a waste of energy to constantly correct. The "bad" and "ugly" communities do not face this pervasive hostility and do not need to engage in these epistemic practices of withdrawal in order to be able to produce and share their knowledge.

The second difference between marginalized communities and the other communities Gaitán Torres considers is that standpoint theory shows that marginalized communities are more likely to produce reliable knowledge than communities about many topics. The

feminist veritism that I argue for in the book takes truth as a valuable epistemic goal. Thus, the fact that marginalized communities are likely to produce truths marks another important difference between the "good" epistemic communities and the "bad" and "ugly" communities. The insights of standpoint theory do not predict that the anti-vaxxer mom group or the group targeting oppressed people will be epistemically reliable.

Finally, I want to say why I disagree with the two criteria Gaitán Torres proposes for distinguishing between good epistemic communities and harmful ones. First, he proposes that fostering negative attitudes and emotions towards other groups is always epistemically harmful. For Gaitán Torres, the reason why negative attitudes are harmful is that they hinder interaction and debate between groups. However as I have argued, interaction and debate between groups is not always epistemically helpful. For privileged people who are often subject to socially situated ignorance, interaction with people different than them can help them unlearn their ignorance. But for marginalized people who are already more likely to have knowledge about the privileged worldview, interaction often systematically undermines and silences this knowledge. Gaitán Torres does not say exactly what he takes to be negative attitudes and emotions, but I think that many examples of what might be taken as such can be epistemically helpful. For example, believing that white supremacists hold a false ideology is a true belief, and anger at white supremacy can motivate people of color to express their knowledge of racist wrongs and also defend themselves from racist attacks (cf. Cherry 2021). Second, Gaitán Torres suggests that good epistemic communities of marginalized communities must "limit themselves to the articulations of experiences of oppression, avoiding wider political issues or debates" (Gaitán Torres 2024). I do not think this is a tenable distinction because it does not recognize the pervasiveness of oppression or how many features of reality are shaped by oppression. Much of what mainstream epistemology takes to be political issues or debates unrelated to oppression are in fact deeply shaped by oppression. For example, in the United States, it is hard to find any wider political issue or debate that is not shaped by the legacies of white supremacy. Gaitán Torres' proposal fails to recognize the scope of the epistemic value of marginalized communities. Conversations between people facing oppression are a critical step in the consciousness-raising process that is central to the production of knowledge recognized by standpoint theory. This process involves articulating one's experiences with oppression, sharing them with others who have similar experiences, and jointly critically reflecting on what these experiences teach us about ourselves, others, political systems, the nature of knowledge production, and many other features of reality relevant to other topics. Thus, to require that marginalized people only articulate their experiences of oppression with each other but refrain from discussing other controversial topics would not only hamper their ability to gain critical insights about many issues, but would also deprive the wider community of these insights.

References

Brinkmann, Matthias. 2022. "In Defence of Non-Ideal Political Deference." *Episteme* 19 (2): 264–85. https://doi.org/10.1017/epi.2020.26.

Cherry, Myisha V. 2021. *The Case for Rage: Why Anger Is Essential to Anti-Racist Struggle*. New York: Oxford University Press.

Citron, Danielle Keats. 2022. *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age*. New York: W.W. Norton & Company, Inc.

Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–57. https://doi.org/10.1111/j.1527-2001.2011.01177.x.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and Ethics in Knowing*. New York: Oxford University Press.

Frost-Arnold, Karen. 2014. "Imposters, Tricksters, and Trustworthiness as an Epistemic Virtue." *Hypatia* 29 (4): 790–807. https://doi.org/10.1111/hypa.12107.

— 2018. "Wikipedia." In *Routledge Handbook of Applied Epistemology*, edited by David Coady and James Chase, 28–40. New York: Routledge.

— 2023. *Who Should We Be Online? A Social Epistemology for the Internet*. New York, NY: Oxford University Press.

Gaitán Torres, Antonio. 2024. "Epistemic Communities and Trust in Digital Contexts." *Daimon: The International Journal of Philosophy*.

Jones, Karen. 2002. "The Politics of Credibility." In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by Louise Antony and Charlotte Witt, 154–76. Boulder, CO: Westview Press.

Levy, Neil. 2019. "Due Deference to Denialism: Explaining Ordinary People's Rejection of Established Scientific Findings." *Synthese* 196 (1): 313–27. https://doi.org/10.1007/s11229-017-1477-x.

McKinnon, Rachel. 2017. "Allies Behaving Badly: Gaslighting as Epistemic Injustice." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus Jr., 167–74. New York: Routledge.

Medina Vizuete, Lola, and Daniel Barbarrusa. Forthcoming. "Am I Still Young at 20? Online Bubbles for Epistemic Activism." *Topoi*.

Mills, Charles. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*, edited by Nancy Tuana and Shannon Sullivan, 26–31. Albany, NY: SUNY Press.

Mitova, Veli. 2020. "Decolonising Knowledge Here and Now." *Philosophical Papers* 49 (2): 191–212. https://doi.org/10.1080/05568641.2020.1779606.

Nagel, Jennifer. 2017. "Comments on The Internet of Us: Knowing More and Understanding Less in the Age of Big Data." Presented at the APA Pacific Division Meeting, Seattle, Washington.

Pohlhaus Jr., Gaile. 2012. "Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance." *Hypatia* 27 (4): 715–35. https://doi.org/10.1111/j.1527-2001.2011.01222.x.

Ridder, Jeroen de. 2022. "Online Illusions of Understanding." *Social Epistemology*, 1–16. https://doi.org/10.1080/02691728.2022.2151331.

Tobi, Abraham T. 2020. "Towards A Plausible Account of Epistemic Decolonisation." *Philosophical Papers* 49 (2): 253–78. https://doi.org/10.1080/05568641.2020.1779602.

RESEÑAS

Daimon. Revista Internacional de Filosofía, nº 93 (2024), pp. 199-236
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico)

HERRERA GUEVARA, A. (2020), *Bioética postsecular e interespecífica: ciencia, ética y cultura en el siglo XXI*, Madrid/Oviedo, Catarata.

Profesora titular de Filosofía Moral en la Universidad de Oviedo, Asunción Herrera Guevara acumula una amplia trayectoria de producción académica marcada por su original reinterpretación de nociones éticas tan centrales a la tradición moderna como la autonomía, el progreso, la ilustración, o la libertad, con la intención de anclar y ajustar a las mismas en un presente cada vez más cambiante y complejo. Si ya su anterior obra, *La Conspiración de la Ignorancia* (2018), reflejaba los recientes virajes de la autora hacia territorios que se extienden más allá de la filosofía moral tradicional, esta última obra puede concebirse como la culminación de su pensamiento en torno a la bioética, esa disciplina actual a la que ya lleva dedicado largo tiempo de su investigación. El presente libro mantiene, pues, continuidad con muchas de las ideas centrales del pensamiento de Herrera, pero se distingue de los anteriores al centrar todo su protagonismo en la bioética y en tratar de ofrecer al público un acercamiento accesible y novedoso a la misma.

El presente es un libro de construir puentes: puentes entre un público lego y los expertos en filosofía moral, entre teoría y práctica, entre ciencias y humanidades y entre lo que dicta la tradición de la disciplina y la mirada siempre crítica y personal de la autora. El primer argumento central del libro consiste en subrayar la paradoja que resulta de combinar la ubicuidad de la bioética como disciplina presente en todos los ámbitos de la vida moderna con la alarmante ausencia de una ciudadanía formada en la reflexión y el método bioéticos. Para combatir dicha paradoja, la autora propone retroceder y profundizar en los cimientos teóricos de la bioética, a saber, a su naturaleza reflexiva, a su método deliberativo y a su dimensión ética.

Para ello, el Capítulo 1 comienza con una redefinición de la bioética que enfatiza en su dimensión filosófica y, en concreto, ética. Poner la ética en el centro de la discusión acerca de la bioética dota a esta última de un carácter epistémico: a saber, hacer bioética requiere saber razonar y formarse en los métodos que ayuden a argumentar, pensar y deliberar sobre las distintas ideas de justicia. Para ello es necesario una iniciación en ética. Herrera propone una serie de definiciones que ayudan a distinguir la ética de otras disciplinas con las que comúnmente se tiende a asociarla. Lo que caracteriza a la ética es que esta tiene un carácter ineludiblemente público y universal, donde la discusión se remite no solo una idea determinada e individual de bien, sino a "una idea de bien para todos por igual, es decir, a la más justa" (Herrera, 2020, 26).

Otro aspecto fundamental de la obra, y que diferencia a Herrera de filósofos como Jürgen Habermas, es la aceptación optimista de la naturaleza doble de la ética, en la que teoría y práctica, normas y valores, principios y contexto se complementan en una visión integradora que permite simultánea-

mente situar a la ética en los contextos reales sin tener que renunciar por ello a su valor normativo. Esto es también especialmente relevante en su visión de la bioética, que la autora define como una disciplina doblemente doble. A su vez, el acercamiento a lo fáctico consiste en un movimiento metodológico que la autora refiere como "naturalismo blando", término tomado del filósofo alemán que Herrera se apropia para vertebrar su propio modelo de método bioético. Tomando su nombre de la propuesta habermasiana por constelar a Darwin con Kant (en Herrera, 2020, 39), el naturalismo blando al que se refiere la autora consiste, pues, en dos tipos de alianzas. La primera hace referencia a la ya mencionada fusión de teoría y práctica. Y la segunda, en la que la autora ahonda en posteriores capítulos, señala la interrelación necesaria entre ciencias y humanidades. En otras palabras, a la necesidad de tender puentes y alimentar la reflexión bioética con los conocimientos aportados por ciencias contemporáneas como la etología o la embriología.

El Capítulo 2 vuelve su interés al significado de la bioética. La anterior es una disciplina transversal en al menos dos sentidos relevantes. Uno, porque está presente en todo tipo de disciplinas, desde la ingeniería hasta la medicina. Segundo, porque afecta tanto a decisiones públicas como privadas. Debido a esto mismo, la autora subraya la necesidad de educar a los ciudadanos en un método y lenguaje que los preparen para afrontar adecuadamente las decisiones y conflictos bioéticos. Una idea constante en la obra es que la bioética habría de situarse firmemente en el marco de la esfera pública universal, y que sus discusiones deberían centrarse en encontrar una idea de justicia que se adecue a "todos por igual". La tarea del ciudadano consiste en formarse para la deliberación pública y para ello es necesario educar en valores cívicos, aprender a separar el lenguaje privado del de los derechos, y fomentar la capacidad de reflexión moral en cada uno.

A pesar de pertenecer a la ética, la bioética se distingue sin embargo por ir más allá de la anterior por poseer una metodología propia y única que, para la autora, la hace más amplia que su antecesora. Mientras que la ética limita el marco de sus discusiones al ámbito humano, la bioética amplía la discusión moral al discutir directamente sobre dilemas medioambientales o en relación con el animalismo. De este modo, la definición e interpretación de la bioética por parte de Herrera se distingue por su carácter especialmente amplio y global, y su salida de lo que llama "el marco antropocéntrico" será también una de las características fundamentales y vertebradoras de todo el libro. El capítulo finaliza con la propuesta de una metodología bioética basada en la conjunción no necesariamente ordenada de tres momentos: el naturalismo blando, referido en el capítulo primero; la corrección normativa, que resalta la necesidad de centrar los debates y dilemas en torno a una idea razonable de justicia; y el momento consecuencialista, que ayuda a facilitar decisiones allí donde los principios tienden a entrar en conflicto entre sí.

En el Capítulo 3, Herrera revisa de forma crítica las dos metodologías más extendidas en la disciplina bioética: el principialismo y el casuismo. A través de un repaso a sus principales representantes, la autora muestra de forma excelente que ni el principialismo ni el casuismo han de entenderse como categorías férreamente opuestas, sino que más bien, un análisis cercano de cada uno muestra como las barreras entre principios y valores, normas y contexto se difuminan en ambos casos. El contexto práctico juega un papel fundamental en modelos de auto-

res principialistas tan fundamentales como Beauchamp, Childress o el español Diego Gracia. Por su parte, también casuistas como Jonsen y Toulmin acuden inevitablemente a los principios bioéticos a la hora de analizar casos. La conclusión de Herrera es pues que, más que como éticas distintas, habría que entender a ambos modelos tradicionales como simplemente diferentes estrategias para afrontar la metodología bioética.

La autora ofrece a continuación su propia propuesta personal de un método bioético basado en dos estrategias fundamentales: la constelación de normas y valores (situándose así en una oportuna posición intermedia entre casuismo y principialismo) y la importancia de introducir la unanimidad como una condición en el debate y la argumentación bioética. Este segundo aspecto, frecuentemente olvidado en las teorías tradicionales, apunta directamente a la necesidad de tener en cuenta las voces de todos los implicados en los dilemas morales. Para garantizar este paso, no obstante, resulta fundamentar establecer unos principios mínimos de justicia que funcionen como el límite inviolable de toda discusión moral. Inspirada en la tradición ilustrada de la que es tan conocedora, Herrera ofrece dos tipos de argumentos convincentes para encontrar las respuestas dentro de esta tradición. En concreto dos principios, el de autonomía y el de libertad (esta entendida como no dominación) son rescatables en tanto en cuanto son resultado de un proceso deliberativo histórico que ha ido ampliando con los siglos el marco de la comunidad moral. Además, ambos son principios multiculturales, susceptibles por tanto de ocupar el centro de la propuesta bioética de la autora.

Quizás la aportación más personal del capítulo esté no obstante en su aportación a las discusiones sobre el estatus de la macrobioética. Entre las diferentes versiones posibles, Herrera aboga por ampliar los extremos de esta última para incluir, diferenciándola así de otras éticas como la ética práctica, las perspectivas ecologista y animalista en su seno. El argumento central de la autora en este punto es de corte kantiano: no existen motivos coherentes, ni desde el punto de vista ético ni desde el científico, para seguir relegando a los animales no humanos de la comunidad moral. Apoyándose tanto en los descubrimientos científicos actuales como en un "ecofeminismo ilustrado" que toma de la filósofa argentina Alicia Puleo, Herrera defiende la necesidad de redefinir los mínimos de justicia de modo que estos se expandan más allá de la lógica antropocentrista e instrumental para incluir a todos los animales sintientes como miembros legítimos de la discusión moral. Este paso no solo surge naturalmente de su método propuesto, sino que también tiene continuidad con propuestas anteriores de su trayectoria académica, en las que la autora sugería ampliar el proyecto emancipatorio de la modernidad con aun una "tercera ilustración" (Herrera, 2014) anclada en un "retroceso sustentable" (Herrera, 2018).

El Capítulo 4 ahonda en la propuesta bioética de la autora al tiempo que trata de anclarla en los retos y circunstancias de nuestro presente. En concreto, Herrera se centra en analizar dos de las características definitorias de nuestra sociedad, a saber, su carácter postsecular tal y como ha sido definido por el filósofo alemán, Jürgen Habermas, y su naturaleza de riesgo global asociado a los avances tecnológicos y neocapitalistas. En primer lugar, Herrera critica ingeniosamente la interpretación de Habermas acerca del significado de lo 'postsecular' para proponer en su lugar una estrategia alternativa e intermedia consistente en constelar los valores personales con las razones políticas que rigen su modelo de ética ciu-

dadana. En otras palabras, lo ideal sería contemplar la postsecularización de la sociedad moderna como una feliz continuidad con la tradición ilustrada, y aprovechar este hecho para consolidar una esfera civil formada en el diálogo político y el razonamiento ético. Esto no quiere decir, para Herrera, que las narrativas personales no puedan, o de hecho no entren a menudo en los razonamientos morales. La idea apunta más bien a la necesidad de conservar estas cualidades y no perder de vista la búsqueda de la justicia universal al mismo tiempo.

En segundo lugar, Herrera reflexiona sobre el significado de vivir en sociedades denominadas de "riesgo global". A menudo dicho adjetivo se asocia exclusivamente al progreso económico y técnico. La lógica economicista y neocapitalista conlleva que los riesgos globales sean asumidos con consecuencias desigualmente perjudiciales y sin el consenso de todos los afectados. Partiendo del esquema ofrecido por Nancy Fraser en *Escalas de Justicia* (2010), Herrera nos recuerda la importancia de asumir un marco de reflexión verdaderamente global, que ella propone ampliar incluso más allá del principio fraseriano de "todos los sujetos" hacia su propuesta no especista por un "principio de todos los seres sintientes". El principio de todos los seres sintientes integra, pues, los modelos de justicia tradicionales dentro de un marco macrobioético que da cabida a las nociones de justicia interespecífica, transcultural e intergeneracional.

El Capítulo 5, el último del libro, está dedicado a la discusión, por parte de la autora, de distintas problemáticas bioéticas siguiendo el método hasta ahora presentado. Lo original de este punto es que los dilemas tratados provienen de distintas narrativas, en concreto, tanto literarias como fílmicas. Con esto, Herrera ahonda una vez más, como es característico en ella, en su sensibilidad con los distintos tipos de relatos para ofrecer la oportunidad para la reflexión ética, un punto en el que también encuentra inspiración en la autora norteamericana Martha Nussbaum. Herrera analiza en concreto cuatro dilemas bioéticos (más precisamente, de la macrobioética). El primer dilema abordado pertenece a la ética ecológica. Herrera propone la película "La Selva Esmeralda" (1985) como un buen ejemplo para ilustrar, a la vez que cuestionar, aspectos problemáticos en torno a nuestro modo de relación actual con la naturaleza, que la autora propone contrastar con los conocimientos científicos actuales, siguiendo así el primer paso de su método, correspondiente al naturalismo blando. El segundo paso necesario en la reflexión bioética, a saber, la reflexión desde unos mínimos de justicia, también es algo que el mero visionado de la película facilita, pues en ella se muestran directa y claramente los daños causados por el antropocentrismo desde el punto de vista de los animales no humanos y la naturaleza.

El segundo dilema pertenece a la filosofía animalista. Herrera profundiza en la necesidad de reflexionar, repensar y reimaginar el mundo desde una conciencia animalista necesaria para su bioética interespecífica, a saber, una conciencia basada en el no antropocentrismo y la no instrumentalización de los animales no humanos. La conciencia animalista es aquella que es capaz de ver más allá de los relatos del "sufrimiento innecesario" y que sabe equiparar el valor de seres humanos y resto de animales sintientes (Herrera, 2020, 142). Llegar a poseer esta conciencia no es sin embargo fácil y requiere un proceso activo de reflexión y entrenamiento en una forma de pensar y sentir contra la *Sittlichkeit* contemporánea. Dada la dificultad de este proceso, Herrera propone dos diferentes tipos de relatos (el libro *La Raza Futura*, de Edward Bulwer-Lytton, 1871, y la película "Mi vida como

un perro", de 1985) para facilitar la entrada a estos dilemas desde un acercamiento situado y encarnado.

El tercer dilema abordado afronta la problemática inminente de la reproducción clónica, que Herrera relaciona ingeniosamente con relatos como el clásico de Mary Shelley, *Frankenstein* (1818), o la película "Código 46", de 2003. Tal problemática, que Herrera interpreta como una nueva manifestación contemporánea de eugenesia liberal fundada en un pensamiento libertario y erróneo desde el planteamiento ético, requiere de un método de reflexión adecuado para hacer frente a sus contradicciones. En este caso, las vivencias en primera persona de los protagonistas de ambos relatos pueden ilustrar de cerca sobre los peligros del naturalismo fuerte y del culto al progreso científico sobre los que la autora nos alerta. Partir desde ambos relatos nos podría llevar, además, a la comprensión no solo sobre las consecuencias indeseables de estas prácticas, sino también a una reflexión sobre los mínimos de justicia que la autora propone a lo largo del libro.

Por último, la película "Sin límites" (2011) ayudaría a abordar problemáticas asociadas a la neuroética contemporánea. El visionado de esta película invita a una reflexión que la autora quiere resaltar: "no existe ninguna vinculación entre capacidad cognitiva y moralidad (Herrera, 2020, 158)." Dicha reflexión resulta importante a la hora de cuestionar determinados prejuicios arraigados (como aquellos que tienen que ver con a qué tipo de sujetos y de prácticas tendemos paradójicamente a considerar inmorales), sino que también reitera la necesidad de reflexionar acerca de ciertos principios. En este caso, Herrera enfatiza la importancia del último paso de su método, a saber, la reflexión y análisis de las consecuencias del dilema y el anclar dichas reflexiones tanto en el ámbito privado como en el público. Vuelve aquí la importancia de anclar la reflexión bioética en una lógica que promueva los intereses de todos por igual.

Con el capítulo 5 se pone fin, pues, a la propuesta de la autora por construir puentes. En este caso, el apoyo de las películas y novelas mencionadas intenta consolidar la voluntad de acercar la reflexión bioética a un público no necesariamente experto. El camino hacia la correcta deliberación moral es largo, arduo y deja más preguntas abiertas que soluciones. El presente libro alerta sobre esto, al tiempo que pretende facilitar este proceso. Todo lo demás será tarea de los futuros lectores.

Alicia García Álvarez
(Universidad de Oviedo)

BRONCANO RODRÍGUEZ, F. (2020). *Conocimiento expropiado. Epistemología política en una democracia radical.* Madrid: Akal.

Fernando Broncano presenta su libro *Conocimiento expropiado. Epistemología política en una democracia radical* como un proyecto de difícil adscripción a ninguna disciplina académica. El texto, que se autoidentifica como "una invitación a la epistemología política tanto a lectores familiarizados con la filosofía como a quienes

les preocupa el lugar del conocimiento en la sociedad" (10), se sabe díscolo ante los programas de ambas disciplinas: la epistemología clásica y de la filosofía política.

La declaración temprana de esta dificultad de adscripción a un espacio académico asentado y reconocido puede generar en la audiencia una primera impresión polarizada del texto: o bien requiere una lectura desde la epistemología, o bien desde la filosofía política. Así, las audiencias doctas en debates epistemológicos pueden sentir que añadir a la epistemología la etiqueta de "política" puede suponer alejarse demasiado de la "pureza analítica" (10) que caracteriza su tradición metodológica. Al mismo tiempo, sin embargo, para el público versado en los debates políticos, la apelación a lo epistémico puede resultar impositiva o ajena al debate democrático.

Esta polarización, no obstante, es una falsa ilusión que esconde precisamente el valor genuino del proyecto de Broncano: la reivindicación de la epistemología política como necesaria para la vida colectiva. En otras palabras, el imperativo de reconocer que toda epistemología es epistemología política y que toda política es política epistemológica.

Dicha hipótesis central es defendida por Broncano a lo largo del texto con una gran rigurosidad filosófica, ya que sabe que tiene que hacer frente a dos tradiciones filosóficas para las cuales esta no es una afirmación evidente. De un lado, la epistemología clásica que aún batalla su tránsito hacia la epistemología social, así como el alcance que ese nuevo adjetivo tiene en sus planteamientos y programas (Fricker et al., 2019). Desde el otro polo, la filosofía política que se enfrenta a la abstracción filosófica y a la falta de reconocimiento de las condiciones materiales de los sujetos políticos de la tradición filosófica. Unido a este rigor filosófico, la estructura del texto entraña una gran destreza narrativa. La decisión del autor de organizar la obra en tres secciones diferenciadas, cada una de ellas con sus respectivos capítulos, asiste precisamente en la elaboración y justificación de la hipótesis central, en tanto que las nociones, tesis y debates que presenta en las secciones iniciales se revelan imprescindibles para la construcción de las argumentaciones de los capítulos posteriores.

Esta sistematicidad en la organización del proyecto hace que Broncano dedique la primera sección de su texto (*Epistemologías vulnerables*) a defender y argumentar la centralidad del conocimiento en la estructura de la sociedad, así como de la epistemología como un componente central de la emancipación en el pensamiento moderno (13). Para sustentar dicha centralidad del conocimiento en el espacio político, Broncano elabora una suerte de genealogía revisada de los hitos más relevantes para el pensamiento moderno (capítulo 2), y también identifica una serie de figuras y transformaciones claves en la epistemología (capítulo 3). Los dos capítulos siguientes (4 y 5), que cierran la sección inicial, avanzan a la audiencia dos nociones que serán transcendentales para el desarrollo de la tesis fundamental del texto: en primer lugar, el descubrimiento del carácter social del conocimiento por parte de la tradición analítica de la filosofía; y, en segundo lugar, las consecuencias de abandonar las concepciones individualistas del sujeto: el reconocimiento de que la dependencia epistémica es omnipresente.

Las enseñanzas recopiladas en la primera parte de su obra serán herramientas claves para entender la parte quizás más teórica del libro (*Epistemologías de la resistencia)*, donde el autor se encarga de responder a una pregunta central: "¿por qué invitar a pensar políticamente la epistemología y epistemo-

lógicamente la política?" (21). Simplificadamente, Broncano defenderá que desligar la política de la epistemología (y viceversa) nos impedirá reconocer una serie de problemáticas que se originan precisamente como resultado de las posiciones sociales de poder y dominación. Para ello, y de manera más compleja, el autor se decantará por recorrer una vía negativa hacia el concepto de justicia que le permite explorar al mismo tiempo dos caras de una misma moneda: los posibles daños o agravios epistémicos sufridos por los agentes y comunidades (incluyendo las injusticias epistémicas) (Fricker, 2007) y los modos de resistencia epistémica (Medina, 2013) que luchan contra las ignorancias individuales y colectivas que aquejan a nuestras sociedades.

Tendremos que esperar a la última sección de la obra (*Epistemologías de la democracia*) para encontrar una propuesta positiva del filósofo. Esta última sección, destinada a la elaboración de una alternativa para los problemas e incertidumbres de las sociedades del conocimiento o de la información (ya apuntados en Broncano, 2019), pasa por la respuesta a dos cuestiones. En primer lugar, la defensa y justificación del valor del conocimiento como base nuclear de la agencia (capítulo 9). En esta empresa será crucial retar tanto la concepción únicamente utilitarista del conocimiento, según la cual se entiende el conocimiento como un bien meramente instrumental, como desafiar la visión del conocimiento como un bien individual y no colectivo. En segundo lugar, el autor necesitará resolver la pregunta por la superioridad epistémica de las democracias frente a otros modelos de producción, distribución y uso del conocimiento (capítulo 10). Para ello recuperará nociones clave presentadas a lo largo del texto, tales como la ignorancia, el testimonio, la dependencia o la agencia epistémica.

Los objetivos hasta el momento señalados y explícitamente descritos por el autor para cada una de las secciones del texto parecen suficientemente contundentes para la defensa de la tesis fundamental de la obra, a saber, que "disputar el concepto de conocimiento es disputar la vida misma" (7). Sin embargo, una lectura pormenorizada de la obra nos permite vislumbrar otros objetivos no-declarados explícitamente por el salmantino que ayudan a sustentar su argumentación. Sirva de ejemplo cómo en la sección inicial del texto ("Epistemologías vulnerables"*)* es posible identificar, al menos, dos empresas implícitas en su escritura. De un lado, la extensa narración y recopilación de las tesis y debates fundamentales en la tradición epistemológica que realiza Broncano demuestra el profundo conocimiento y rigurosidad analítica de su evaluación. Este lúcido análisis del filósofo, que claramente no busca la erudición vacía, lo que logra es afianzar el proyecto que constituye *Conocimiento expropiado* ante potenciales críticas respecto a su "pureza analítica". Del otro lado, esta primera sección del texto consigue poner en jaque de manera convincente algunos de los grandes presupuestos sobre los que se asienta la epistemología clásica. De esta forma, afirmaciones transgresoras como la identificación de Hegel como "creador de la epistemología histórica, social y política" (73) ponen de manifiesto su intención de leer la epistemología desde unos presupuestos renovados.

La misma práctica, según la cual el autor entra en diálogo con algo más de lo explícitamente reconocido por su texto, parece estar también detrás de las discusiones que rescata en torno a la superioridad epistémica de la democracia frente a otros modelos políticos y epistémicos. Broncano explícitamente entra a debatir y rebatir las tendencias tecnocráticas y epistocráticas, que persona-

liza en Jason Brennan (369). Para ello utiliza argumentos epistémicos convincentes con los que defiende la superioridad de las democracias frente a otros modelos organizativos de producción y transmisión del conocimiento. Sin embargo, en su argumentación se aprecia una premura por la defensa férrea de la democracia que no llega a estar suficientemente justificada en el texto. La falta de una explicación más clara de esta urgencia, quizás por una cuestión de extensión, quizás por decisión metodológica,[1] ocasiona que la narración pierda la oportunidad de poner en valor precisamente la relevancia de proponer la "epistemología política como la primera línea de resistencia al peligro" (26), un peligro que, no obstante, queda falto de concreción. Así, a pesar de simpatizar con su preocupación y la necesidad de una defensa seria de los modelos democráticos, el texto falla en justificar con claridad que la propuesta de una epistemología política es crucial en el momento político y epistémico *actual*, y por qué. De esta forma, desaprovecha la oportunidad de fortalecer la valía del proyecto, precisamente por no dar razón, en primer lugar, de la deriva política de las democracias actuales, pero, y quizás más importante aún, por no dar cuenta de la deriva intelectual y epistémica de los agentes e instituciones generadoras y distribuidoras de conocimiento.

La pérdida de esta oportunidad empero no oscurece el enorme aporte que la obra supone. La contribución que el texto constituye puede medirse en su capacidad de recoger diestramente los debates de diferentes disciplinas, la originalidad de las hipótesis de trabajo o los avances únicos en debates de plena actualidad en la filosofía. No obstante, no se debería dejar de comprender *Conocimiento expropiado* fundamentalmente como una obra de resistencia, subversión epistémica y democratización del conocimiento para con la comunidad hispanohablante. Acercar de manera tan diestra a la lengua española debates, autores, textos y problemáticas casi exclusivamente accesibles en inglés es en sí mismo una práctica orientada al desmantelamiento de ignorancias e injusticias epistémicas (Fricker, 2007; Medina, 2013; Dotson, 2012) En este sentido son también de alabar los esfuerzos del autor por conectar cada problemática con ejemplos reales, cotidianos y generalmente conocidos por cualquier audiencia, tanto nacional como internacional. Las referencias a redes sociales como Facebook, Twitter o Google, a artículos periodísticos, a declaraciones de presidentes de gobiernos estadounidenses o españoles, a la Guerra del Golfo, a procesos judiciales como los de Plácido Domingo, movimientos sociales como #MeToo o del VHI, o a películas como *Carmen y Lola*, son asideros que permiten aterrizar nociones y conceptos, que de lo contrario podrían resultar inaccesibles o muy distantes de la vida cotidiana.

La obra de Fernando Broncano contribuye así de manera generosa a la elaboración de un panorama general desde el que mirar a la sociedad actual y del que se pueden nutrir múltiples disciplinas, con especial interés para la epistemología y la filosofía política.

1 Broncano sí que recoge preocupaciones de este calado en su libro *Puntos ciegos. Ignorancia pública y conocimiento privado* (2019) con mayor extensión y parsimonia que en el texto que nos ocupa, especialmente en la segunda sección del libro (Túneles de la mente y autopistas de la información). Puede que la decisión de no incluir un debate más sosegado al respecto en este texto responda precisamente a la disponibilidad de ese libro anterior, ya que en un par de ocasiones el propio Broncano nos emplaza a *Puntos ciegos* para completar su argumentación. Sirva como ejemplo la nota 1 del capítulo VIII (Epistemologías de la resistencia).

La asunción de que su público no deba tener conocimientos previos de ninguna disciplina concreta permite al texto avanzar en problemáticas técnicas para la filosofía (problema de los expertos, variedades de injusticias epistémicas, autonomía y agencia…) pero también en debates centrales para la vida de cualquier persona (elección de modelos políticos más adecuados, acceso a la información, privacidad y gobiernos…). El texto se establece de esta manera como un diagnóstico de referencia para las problemáticas e inquietudes de varias disciplinas, pudiendo servir como punto de partida para otras novedosas y profundas investigaciones futuras.

Referencias

Broncano, F. (2019). *Puntos ciegos: ignorancia pública y conocimiento privado*. Lengua de Trapo.

Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression. *Frontiers: A Journal of Women Studies*, *33*(1), 24-47.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Fricker, M., Graham, P. J., Henderson, D., & Pedersen, N. J. (Eds.). (2019). *The Routledge handbook of social epistemology*. Routledge.

Kidd, I. J., Medina, J., & Pohlhaus Jr, G. (Eds.). (2017). *The Routledge Handbook of Epistemic Injustice*. Taylor & Francis.

Medina, J. (2012). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press.

Lola Medina Vizuete[2]
(Universidad de Sevilla)

2 (lmvizuete@us.es). Esta investigación se ha financiado con cargo al proyecto E-RISK, "New Perspectives on Epistemic Risk" (PGC2018-098805-B-I00), MCIU/AEI/FEDER, UE.

NEGRI, A. (2021). *Spinoza ayer y hoy*. Buenos Aires: Editorial Cactus.

El ensayo de Antonio Negri *Spinoza ayer y hoy* es el cuarto escrito dedicado al filósofo herético y forma parte de una compleja constelación del pensar spinoziano[1]. Además, es el último escrito dirigido a analizar las transformaciones en su pensamiento (*Marx y Foucault* y *De la fábrica a la metrópolis*). Esta investigación está constituida por tres partes: la primera 'Spinoza en el 68', muestra *un buen momento* para afirmar un pensamiento democrático en la lucha de clases. Es decir, la coyuntura de los años sesenta es crucial para el surgimiento de nuevas interpretaciones spinozianas. La segunda parte 'Spinoza hoy' se hace cargo de problemas conceptuales spinozianos: la potencia del actuar, libertad, multitud e inmaterialidad, especialmente en la época

1 *Spinoza ayer y hoy* se articula sobre la base de tres obras: *La anomalía salvaje* (1981), *Spinoza subversivo* (1992) y *Spinoza y nosotros* (2010).

del capitalismo neoliberal. La última parte "Spinoza en el siglo XVII", proporciona los elementos para pensar la filosofía política de la modernidad desde Descartes a Spinoza.

No parece casual que el pensamiento de Negri termine estas reflexiones con un ensayo sobre Spinoza, pues el análisis del filósofo italiano requiere no solo de la teoría, sino también de la estrategia para la lucha. Así, Spinoza no está entregado a una interpretación etológica, sino que se presenta como una ontología constitutiva. Negri menciona que, en Deleuze y Guattari, Spinoza se convierte en un revolucionario. Deleuze en *Diferencia y repetición* (2012), un texto escrito en un *tono bergsoniano*, plantea una ontología virtual, es decir, la ontología de la posibilidad. Sin embargo, en la filosofía de Spinoza Deleuze encuentra un pensamiento en acto, en movimiento, no solamente una virtualidad. En el *Anti-Edipo* (2010) las máquinas deseantes se presentan como máquinas sociales. Deleuze y Guattari suponen que la producción del deseo existe desde el inicio, entonces, hay producción del deseo desde que surge la producción y reproducción social. Las máquinas son producción de deseos. El *Cuerpo sin Órganos* está constituido por un deseo descodificado que se da en un plano de inmanencia como una producción de subjetividades deseantes. En Spinoza la razón y la acción no aparecen separadas. "Cuerpo y Mente están absolutamente asimilados, allí donde el ser ya no es orden sino producción de orden, multiplicidad de modos, fusión de acontecimientos, es decir, potencia" (p. 57). Para Deleuze, Spinoza significa la producción social desde el trabajo de las singularidades.

En 1968 Deleuze y Matheron recuperan a Spinoza como un proyecto de inmanencia, pensándola como una singularidad activa. Asimismo, la segunda generación de lectores de Spinoza que va de Balibar a Moreau comienzan a plantear el problema del movimiento de la inmanencia, es decir, los espacios productivos donde se mueven las singularidades. Moreau supone que, "el plano de inmanencia es completamente rearticulado como tejido ontológico de singularidades" (p. 67). En efecto, la inmanencia y su movimiento desactivan cualquier intento de un orden político trascendente. Laurent Bove en su interpretación de Spinoza considera que la 'estrategia del conatus' se basa en una proyección infinita de la existencia de la singularidad. "Es una máquina de esencias y de eternidad" (p. 70). La lectura de Bove supone que el *conatus* en su movimiento deviene imaginación convirtiendo la singularidad en una que es: libre, activa y constitutiva. En este sentido, el *conatus* también sería *resistencia*, pues en su movimiento evita la tristeza y el odio, es decir, se convierte en una esencia de amor y alegría en la lucha. El *conatus* reclama su orden constitutivo en la organización de la vida política, donde produce un acto de liberación no solo individual sino colectivo. El *conatus* entonces rearticula la política como un asunto de la colectividad.

La segunda sección "Spinoza hoy" está dedicada a problematizar la actualidad conceptual de Spinoza. La interpretación de Negri sobre la filosofía spinoziana se caracteriza por una potencia ontológica, inmanencia política y monismo estratégico, es lo que denomina: *otra potencia*. "El pensamiento de la potencia como constitución no insiste tanto en la marca de la individualidad como en la de la singularidad modal" (p. 76). Es una potencia que se organiza de forma modal, común y horizontalmente. La potencia spinoziana no es una *potentia vis individua,* es decir, en el ontologismo spinoziano se elimina cualquier forma de trascendentalismo que pretende acumular la potencia de los individuos. En la potencia spinoziana no

existe ninguna teleología, sino que está atravesada por el conflicto de las singularidades. En términos políticos la simetría entre *potentia* y *potestas* se desarticula por una sobreabundancia de la *potentia* sobre la base de la *cupiditas*. Es decir, la *potentia* posee la fuerza productiva del amor. La *cupiditas* (deseo) provoca una asimetría entre la ontología constituyente y el poder constituido. Finalmente, en Spinoza la potencia es una fuente de ruptura constitutiva que excede toda medida. "Solo hay potencia, es decir, libertad que se opone a la soledad y que construye lo común" (p. 81). La *potentia* es una producción continúa hacia lo común, como un excedente que continúa construyéndose. Esta interpretación de Spinoza nos conduce a un pensamiento de singularidades constituyentes que son fuente inagotable del orden jurídico y político.

Así pues, el concepto de multitud como algo constitutivo se basa en una reinterpretación del término democracia. En el *Tratado teológico-político* Spinoza concibe la democracia de los hebreos como una realidad ético-política que pretende eliminar todo rastro de trascendencia de la Ley. "En el *TP*, en cambio, el concepto de democracia está totalmente secularizado" (p. 91). Ciertamente, los elementos trascendentales de la transferencia de poder son eliminados. La legitimidad del poder político no depende de la transferencia de poder al soberano. Democracia es *omnino absolutum imperium* que se presenta como potencia de los sujetos, articulación siempre abierta de la vida. La multitud es el movimiento incesante de la potencia de las singularidades. "La *multitudo* no es algo orgánico, no es un Uno sino una multiplicidad" (p. 94). Negri dice que la multitud es una máquina de producción de subjetivación infinita. "Multitud es, por el contrario, aquella de los hombres que, en estos mismos días, atraviesan Europa buscando trabajo y felicidad huyendo de ese universo de pasiones tristes, de supersticiones asesinas y de odiosas humillaciones en el que estaban encerrados" (p. 98). Son singularidades que participan conjuntamente en la producción de lo común.

La producción del *conatus* —diría Laurent Bove— pretende resistir cualquier pasión que obstruya su constitución. El odio en Spinoza es aquella pasión que obstruye el desarrollo de la potencia, "el odio no es otra cosa que la tristeza acompañada de la idea de una causa exterior" (p. 173). En la *Ética* Spinoza menciona que el alma se imagina aquellas cosas que favorecen o aumentan la potencia del actuar, el odio bloquea la potencia de la acción. Para las singularidades, el mundo en un sentido colectivo debe ser la institución del amor, cuya vida en el trabajo es la construcción del común, es un esfuerzo por salir de la soledad. Entonces, el *conatus* se desarrolla como *cupiditas*. El proletariado odia la explotación de clase, pero "construye comunidad y organización a través de una práctica de cooperación en el trabajo" (p. 182). Negri sostiene que el proletariado invierte el odio en una fuerza productiva de resistencia, cooperación y construcción del común. La acción del proletariado deviene la *práctica del amor*.

Finalmente, la última sección enfatiza el pensamiento político en la modernidad. En las antípodas de la modernidad y el germen del desarrollo capitalista, existe un desarrollo de la técnica y de las nuevas fuerzas productivas que fueron sujetadas a nuevas formas de control. En ese momento se pretende sustraer toda potencia singular, es decir, se drena cualquier posibilidad de rebelión. "La expropiación de lo común, tal como se desarrolló en el proceso de acumulación originaria, es transfigurada y, por lo tanto, mistificada en la invención de la utilidad pública" (p. 189). La inmanencia

pretende invertir esa lógica, significa que no hay *afuera* de este mundo, sino que la posibilidad de vivir se encuentra *adentro*. "En Spinoza, las fuerzas productivas producen las relaciones de producción" (p. 191). La construcción de la subjetividad se articula desde las singularidades, que no apelan a ningún *afuera*, sino que son una *ontología de la actualidad* producida por la multiplicidad de sujetos. "Se afirma una ontología de la actualidad en el momento en que las subjetividades producen y se constituyen en lo común" (p. 199). Entonces, una política de la inmanencia es la negación de un poder trascendente, es el trabajo constituyente de las singularidades de lo común.

El filósofo italiano considera que la política no se puede separar de su expresión metafísica, la historiografía debe de tomar en cuenta estos supuestos, es decir, no debe de pretender renunciar a estudiar la totalidad de las articulaciones. Ciertamente, el Estado no puede estudiarse sin verificar su relación con la sociedad civil. ¿Cómo articular un estudio del Estado? Lo que caracteriza el surgimiento del Estado moderno es: "El poder del soberano y la articulación mecánica y racional de la expresión de su voluntad a través de un aparato de poder" (p. 222). Chabod piensa que lo que caracteriza al Estado moderno en su nacimiento es su nueva organización. En efecto, el 'estado máquina' se organiza en torno a una idea de orden social, económico y religioso. "Chabod ratificaba […] que el proceso de constitución, del Estado moderno y la sociedad civil tuvo inicio en el siglo XIV en Italia" (p. 230). En este sentido, el humanismo renacentista puso en primer plano los valores de la libertad y autonomía. Sin embargo, la naciente burguesía descubre la necesidad del orden social, planteándose el problema del desarrollo de la libertad y el orden.

Como se ha dicho: la burguesía surgió de la crisis renacentista como clase socialmente hegemónica. Pero aquí debe subrayarse la importancia de dicha afirmación: Sin tener la posibilidad de conquistar políticamente el Estado, la burguesía, con su presencia social simple, masiva y decisiva, lo configura a su imagen y semejanza. El Estado es una máquina para mantener el orden: social, económico y religioso. Pero la sociedad, la economía y la religión están dominadas por el sentimiento y por la acción burguesas (p. 252-53).

En la construcción del Estado máquina existe una disociación entre el Estado y la sociedad civil. Es el momento en el cual la burguesía surge como clase política separada. Así, la burguesía como una clase social libre puede participar en el proceso de acumulación primitiva y se erige como la clase política dominante, para posteriormente erosionar progresivamente al Estado máquina.

Este ensayo sobre Spinoza pretende construir una ontología productiva, un pensamiento que suspenda todo indicio de trascendencia política. A través de Spinoza, el filósofo italiano pretende que la metafísica se convierta en el lugar donde emergen las singularidades activas. Debido a que la metafísica es una política de lo concreto, ahí se presenta la posibilidad y el antagonismo. Es decir, la metafísica spinoziana que plantea Negri va en dirección opuesta a la obediencia y la trascendencia normativa. La metafísica spinoziana es construcción activa de las singularidades, es un antagonismo y una refundación del poder político constituido.

Luis Alberto Jiménez Morales
(Universidad Autónoma Metropolitana Xochimilco)

LOUGHLIN, M. (2022). *Against Constitutionalism*. Cambridge, Massachusetts: Harvard University Press.

La reciente sentencia del Tribunal Supremo norteamericano estableciendo que el lugar para decidir sobre la despenalización del aborto no es la Constitución, sino los poderes legislativos de cada Estado, ha puesto sobre el tapete un conflicto larvado en las últimas décadas entre el constitucionalismo y la democracia constitucional. Lo que es aparentemente paradójico, pero al menos esta es la tesis de Martin Loughlin en *Against Constitutionalism*, que ilumina algunas cuestiones básicas sobre la confusión actual que sufrimos entre populismo, globalismo, democracias iliberales y desconfianza generalizada en las instituciones políticas, las élites sociales y el futuro de lo que los más optimistas e ingenuos, valga la redundancia, denominaron "el fin de la Historia" (Francis Fukuyama) o "el ocaso de las ideologías" (Lucio Colletti).

Para empezar, ¿qué entiende Loughlin por "constitucionalismo"? No es, como podría parecer en primer vistazo, sinónimo de una actitud favorable a la democracia constitucional o liberal, como prefiera llamarse. En realidad, y esta es la tesis fuerte de Loughlin, es justamente el constitucionalismo una filosofía del gobierno que se ha convertido en la más influyente ahora mismo, pero que subvierte y anula (transforma, en sentido de destrucción), la concepción clásica de democracia constitucional.

¿Qué es lo que pretende el constitucionalismo así entendido? Cuanta más guía constitucional de las sociedades, sostiene Loughlin, más influencia de las élites que están al mando de las instituciones y las diseñan, y menos poder para el pueblo, al que solo le queda obedecer lo que dichas élites institucionales les explican condescendientemente y obligan coactivamente. En la democracia constitucional clásica, las instituciones y los representantes canalizan el poder y la voluntad del pueblo. Sin embargo, el constitucionalismo es una especie de reverso tenebroso de la democracia constitucional, ya que aunque formalmente parecería ser su expresión más acabada, en realidad, como decía, trastoca su funcionamiento, anulando la parte democrática y convirtiéndose el constitucionalismo en un eufemismo de una versión de la dictadura platónica de los filósofos-sabios.

Esta es, en suma, la tesis fuerte de Loughlin (p. x), que el «el constitucionalismo es una forma aberrante de gobernar que debe ser superada si se quiere mantener la fe en una democracia constitucional, este libro defiende la democracia constitucional contra el constitucionalismo.»

La Constitución, entendida a la clásica manera del siglo XVII y XVIII, es un documento escrito por la élite, pero siempre en nombre del pueblo, que establece cuáles son los poderes del gobierno, el modo de llevarlos a cabo, los derechos básicos de los ciudadanos, y regula las relaciones entre las instituciones gubernamentales y los ciudadanos. Cada Constitución en cada lugar del mundo debía ser, por tanto, diferente, aunque tuviesen todas ellas un aire de familia. Pero era fundamental que cada una estuviese impregnada de un espíritu de las leyes, que diría Montesquieu, particular de una cultura y propia de una forma de ser. Frente a esta visión relativista (aunque sería más apropiado decir "interculturalista" porque mantiene una dialéctica entre lo particular y lo general), la aproximación constitucio-

nalista es universalista y, en el límite, plantearía una Constitución planetaria, idéntica para todo el mundo, en cualquier cultura, modo de vida, religión mayoritaria o clima. Este conflicto es el núcleo de la paradoja constitucional que subyace a nuestras democracias, de nuevo, según Loughlin (página 7) «este contraste entre el pluralismo del gobierno constitucional y el universalismo del constitucionalismo.»

Esta tendencia hacia la universalización se ha traducido en una santificación de la Constitución, que se ha convertido, como se temía Thomas Jefferson, en algo tan sagrado que muy difícilmente puede ser reformado. En lugar de ser un documento útil para el establecimiento de un marco gubernamental estable, se está convirtiendo, a través de esta filosofía política constitucionalista, en una especie de tótem para su adoración acrítica, con sus artículos convirtiéndose en dogmas que hay que respetar sin debate.

El problema, por tanto, aparece cuando la Constitución pasa de ser un proyecto para crear un órgano de gobierno a convertirse ella misma en un sistema autosostenible de poder, con su sistema de jueces en distintos niveles que llegan a ser no solo intérpretes de la ley, sino un poder legislativo en paralelo al poder de los representantes elegidos por el pueblo. De este modo, los agentes constitucionales, en toda la escala judicial, se conjuran para mantener el orden constitucional en cada momento histórico, haciendo ajustes en las leyes y creando derechos, de manera que se mantenga lo que ellos consideran que es el equilibrio del sistema de la Ley Fundamental.

Hay dos hitos históricos, según Loughlin, en este tránsito desde una democracia constitucional a un sistema constitucionalizado. En primer lugar, la guerra de Secesión. Más tarde, el New Deal de Roosevelt. La guerra norteamericana de finales del XIX puso en cuestión el proyecto de república única que se había fundado sobre el hecho de la conquista y consolidado a través de la institución de la esclavitud (la cual había sido proscrita del Reino Unido por estar en contradicción con su "common law"). Era un paradójico "imperio de la libertad" en palabras de Jefferson, un Padre Fundador que no tenía ningún escrúpulo moral en poseer varios centenares de esclavos.

Toda esta concepción constitucional fue desafiada con la Guerra Civil, que implicaba una reinterpretación no solo de la letra de la Constitución, sino también de su espíritu, avanzando en su formulación como un documento sagrado en el que los derechos se convertían en principios abstractos universales no debatibles. Así, se convirtió la Constitución en un sistema autocontenido y autosuficiente que creó su propia dinámica de poder y sus adherentes sacerdotales, llegando más allá del simple marco para desarrollar la acción política.

El New Deal de Roosevelt llevó a cabo otra vuelta de tuerca en la entronización republicana de la Constitución y del Tribunal Supremo, a pesar de las críticas y las resistencias de primera hora al programa político del presidente, como ejerciente de una tutela vital ante el Tribunal Supremo.

Fuera de los Estados Unidos han sido Alemania y la India las que han llevado a cabo nuevos proyectos constitucionales basados en la filosofía del constitucionalismo. Alemania, a través del diseño de una democracia militante que asegura que el núcleo del régimen político –del sistema federal a la indisolubilidad de la nación, pasando por la protección de los derechos fundamentales– es invulnerable al cambio constitucional.

El caso de la India es también paradigmático, según Loughlin. En un país con tantas religiones, lenguas y diversos tipos de identidades comunitarias, la misión encomendada a la Constitución fue la de "liberar"

a los individuos de dichas comunidades convirtiéndolos en ciudadanos. Además, dado el escaso índice de desarrollo educativo y la presencia de múltiples tradiciones que mantienen la desigualdad y la falta de "simpatía social", la Constitución incorpora una serie de Principios Directivos de la Política de Estado que hacen de ella una Constitución partidista siguiendo las características, establecidas como esenciales, de la democracia, la igualdad, el federalismo, el estado de derecho, el secularismo y el socialismo.

Tanto en Alemania como en la India, salvando las distancias, no existiría la moralidad constitucional básica para que pudiera funcionar un texto constitucional, por lo que la diseñaron, por un lado, como militante y, por otra, se incorporaron multitud de cuestiones administrativas que, en rigor, correspondían al plano legislativo. Esto significa que si en Alemania se convierte la Constitución en un fundamento nacional, dado el terrible pasado del que emerge el país germano tras la Segunda Guerra Mundial, en la India se constitucionaliza tanto en la vida diaria como la civil. De ahí la democracia militante en Alemania y la democracia paternalista en la India.

En cualquier caso, da igual que seas un régimen constitucional nuevo o de asentada trayectoria, lo cierto es que la revisión judicial constitucional de la vida política común ha aumentado exponencialmente, por lo que cada vez hay más protestas sobre la judicialización de la política, por una parte, y acusaciones de "lawfare", por otra, siendo dicho "lawfare" la realización de la política partidista a través de la instrumentalización de los órganos judiciales.

Desde el punto de vista filosófico, Loughlin identifica este constitucionalismo con lo que denomina "la segunda fase de la modernidad", caracterizada por lo que Max Weber denominó "desencantamiento" del mundo a favor de la ciencia, el secularismo, la burocratización y el racionalismo. Lo que lleva, a su vez, a una erosión del poder de los gobiernos y de la soberanía de las naciones-estado ante el triunfo de la globalización, el comercio internacional, la tecnología y las redes de comunicación. Si en la primera fase de la modernidad el sistema político-económico dominante fue el liberalismo del "laissez faire", en esta segunda fase es el neoliberalismo el que domina, con su acento keynesiano en la necesidad de un Estado firme, fuerte y expansivo para hacer que los mercados puedan desarrollarse. Un Estado que, como señalábamos, ya no es el propio de la nación, sino que es fundamentalmente el internacional, como el que corresponde con instituciones al estilo de la ONU, la OMS, el FMI o la UE. Todas ellas instituciones que de alguna manera u otra interfieren, regulan e intervienen en los estados-nación, implementando políticas, ayudando económicamente (con la contrapartida teórica que establezcan dichas instituciones), y estableciendo marcos ideológicos y culturales. A este estado de globalización cosmopolita lo denomina Loughlin "ordo-constitucionalismo".

Llegado a este punto cabe establecer otra definición loughliniana de constitucionalismo (p. 21): «Conjunto de principios que instituyen un orden global fundado en principios bastante abstractos de racionalidad, subsidiariedad y proporcionalidad». Mientras que la democracia constitucional y la democracia de masas son la cara y la cruz de un mismo sistema democrático, el constitucionalismo y la democracia de masas son incompatibles. Y esta última contradicción se explica por el hecho de que el papel de la Constitución ya no es, como en el principio de su trayectoria, un instrumento sin más para la toma de decisiones colectivas, sino que se ha convertido, con el constitucionalismo, en una representación simbólica de una identidad política colectiva concreta.

El libro está dividido en tres partes, dedicadas respectivamente a los orígenes del constitucionalismo, sus elementos y la era del constitucionalismo. Son los epígrafes de esta última sección los que nos dan las claves de por dónde va el libro: *La Constitución como Religión Civil*, *Hacia la Juristocracia*, *Integración a través de la Interpretación* y *Nuevas Especies de la Ley*. Los autores más relevantes con los que discute o en los que se apoya Loughlin desde el punto filosófico son Hans Kelsen, Jürgen Habermas, Carl Schmitt, Max Weber y Friedrich Hayek.

Sírvanos de muestra este último del diálogo que establece Loughlin con todos ellos. Hayek mantiene que en el corazón de la Modernidad hay dos modelos contrapuestos. Por un lado, el racionalismo constructivista que tiene su fundación en la filosofía de Descartes, según la cual la sociedad puede ser diseñada como un ejercicio de la razón consciente. Esta cosmovisión cartesiana es la que estaría, según Hayek, tras la ingeniería social subyacente a todos los gobiernos modernos y, sobre todo, al socialismo. Como derivada de este proyecto racionalista constructivista tendríamos el fenómeno paralelo de la destrucción de la libertad y la servidumbre más o menos voluntaria. El nacionalsocialismo de Hitler y el comunismo de Stalin no serían sino las versiones extremas de este proyecto cartesiano transmutado en ingeniería social total.

El segundo modelo de Modernidad es defendido por Hayek en el tercer volumen de *Ley, Legislación y Libertad*. Hayek es fundamental para Loughlin porque en estas dos obras reivindica la democracia constitucional frente al nuevo constitucionalismo. También porque da pie a la introducción del concepto "Ordoliberalismo", vinculado a la rama alemana de la Sociedad Mont Pelerin, un *think tank* organizado por Hayek tras la Segunda Guerra Mundial para debatir las ideas liberales entre intelectuales simpatizantes del movimiento. Los ordoliberales alemanes, como Walter Eucken, defendían que una Constitución económica era fundamental para que desde un Estado fuerte se defendieran los mercados libres, ya que en caso contrario el *laissez faire* haría que los monopolios y cárteles destruyesen la libre competencia. Hayek establecía así una agenda neoliberal para que la red de instituciones globales fuesen evolucionando para establecer un sistema mundial libre de la interferencia política usual para, finalmente, llegar a lo que denominaba «el destronamiento de la política».

Pero, contra el deseo de Hayek, este destronamiento de la política, entendida esta como proceso democrático vinculado a las tradiciones y al "common law", ha sido hecho a través del entronamiento del texto constitucional concebido dentro de los parámetros del racionalismo constructivista que tanto criticaba el filósofo vienés. ¿Cómo? A través de la obligación constitucional impuesta al legislativo y el ejecutivo para realizar determinados valores específicos. Lo que Loughlin denomina "aspirational constitutions" dentro de un "emancipatory project".

Los nuevos proyectos constitucionales se están desarrollando en la actualidad, como es el caso de Chile, dentro de este nuevo modelo constitucionalista. Lo hacen en un contexto filosófico específico, la segunda modernidad, que establece una dualidad entre el incremento del individualismo, por un lado, y la escala global de las relaciones sociales, culturales y económicas. Esta dualidad ha hecho dinamitar las jerarquías tradicionales y ha disuelto los vínculos comunitarios, lo que ha llevado a considerar la Constitución como la forma social de la nación, con un giro reflexivo por el que los derechos es la principal ocupación constitucional, en lugar de las instituciones, con lo que los jueces toman el lugar de los legisladores, convirtiendo en política los

principios de racionalidad, proporcionalidad y subsidiaridad. Todo este proceso es lo que denomina Loughlin "constitucionalización", uno de cuyos efectos es la difuminación entre lo nacional y lo internacional, como se manifiesta en el concepto de "jurisdicción universal", cuyos principales defensores son jueces, de Benjamin Berell Ferencz, fiscal de Nuremberg que trabajó para el establecimiento de una Corte Penal Internacional, al juez de Senegal Demba Kandji, que en 2000 ordenó la detención del sátrapa africano Hissène Habré. Esta internalización de la política que promueve el paradigma constitucionalista ha implicado el incremento del poder de gobernar de las instituciones internacionales y a través de acuerdos intergubernamentales, fundamentalmente basados en principios universales de razón pública antes que por la decisión de los pueblos. Lo que implica el triunfo del neoliberalismo basado en el cosmopolitismo cultural y los mercados como principio económico.

Comentábamos al inicio cómo el análisis de Loughlin ilumina la polémica surgida a raíz de la sentencia del Tribunal Supremo norteamericano sobre el aborto. Por su parte, Loughlin termina el libro mencionando el surgimiento de partidos populistas en todo el mundo, a izquierda y derecha. Lo que usualmente ha sido interpretado como un desafío a la democracia constitucional. Pero a la luz del análisis de Loughlin la rebelión de estos partidos no es tanto contra dicho tipo de democracia, sino contra su variación en el paradigma constitucionalista (lo que no significa que sean defensores de la democracia constitucional). El populismo tendría así, en la visión de Loughlin, un sentido de la de "reivindicación democrática". del poder constitucional clásico frente a la deriva constitucionalista. Paradójicamente, a más populismo, más reacción constitucionalista, reforzándose ambos movimientos dialécticamente. La navegación constitucional legítima, defiende Loughlin, pasa por la Escila del populismo y la Caribdis del constitucionalismo para restaurar los valores básicos de la democracia constitucional.

Referencias Bibliográficas

COLLETTI, L. (1982). *La superación de la ideología*. Madrid: Cátedra.

FUKUYAMA, F. (1992). *El fin de la historia y el último hombre*. Barcelona: Planeta.

HAYEK, F. (2014). Derecho, legislación y libertad. Madrid: Unión Editorial.

LOUGHLIN, M. (2022). *Against Constitutionalism*. Cambridge, Massachusetts: Harvard University Press.

Santiago Navajas
(Universidad de Granada)

FERNÁNDEZ LÓPEZ, J. A. (2022). *Estudios de pensamiento medieval hispanojudío*. Madrid: Comillas, 196 pp.

Estudios de pensamiento medieval hispanojudío es un recorrido panorámico por algunos de los hitos fundamentales del pensamiento judío medieval en tierras peninsulares. A modo de jalones en un camino histórico y cultural, la historiografía, la teología política, la mística, la Escritura, la Tradición y, de forma privilegiada, la filosofía,

se presentan en esta obra como estímulo intelectual para una sugerente historia cultural del judaísmo. Este trabajo aporta sin duda una mirada renovada sobre una temática de la que no existen en castellano la referencias que merece (*Pensamiento y mística hispano judía y sefardí*, coordinado por Judith Targarona, editado hace más de veinte años; *El legado del judaísmo español*, de David Gonzalo Maeso, un clásico de 1972 concebido como breve manual universitario; *Filosofía medieval hispana*, un conjunto de colaboraciones coordinado por Joaquín Lomba). Las dos primeras no son estrictamente filosóficas, aunque abordan algunos aspectos de la filosofía judía medieval. De las temáticas que aborda el libro, tres de ellas son prácticamente originales en nuestra lengua. Me refiero a los capítulos dedicados a Ibn Daud, a la Cábala, desde una perspectiva académica y el estudio de la traducción vernácula de la *Guía de perplejos*. Los dos restantes, son también dos aproximaciones novedosas al pensamiento judío medieval hispano: una reflexión sobre la profecía y la intelección humana y una estudio sobre la figura de Hasdai Crescas. Como obra de naturaleza filosófica, *Estudios de pensamiento medieval hispanojudío* es una aproximación desde la filosofía a la concepción de la historia, la tradición religiosa, la mística y el pensamiento judío en tierras peninsulares que abarca los siglos XII-XV.

La dialéctica entre la fe y la razón, entre lo sagrado y lo profano, entre el devenir de la historia y la permanencia de la Tradición es el nudo gordiano del que parten todas las manifestaciones morales, espirituales e intelectuales del judaísmo antiguo y medieval. La filosofía religiosa judía representa el fascinante punto de encuentro entre la racionalidad humana y el mundo de la divinidad. Un proceso de comunicación "descendente", que posee como contrapunto, desde finales de la Antigüedad, la fascinante búsqueda de comunicación mística con los secretos escondidos de Dios en sentido "ascendente" y que eclosiona con fuerza en el Medioevo en la cábala peninsular. La religión judía, se verá obligada a replantearse su propia esencia en una permanente dialéctica interna y en conflicto con la crudeza de los acontecimientos históricos, inclusive con vigor renovado en un siglo, el XV, en el que, paradójicamente, culmina de forma trágica la historia del judaísmo hispano.

Recogiendo esta caracterización, el primer capítulo de *Estudios de pensamiento medieval hispanojudío* es una aproximación a estas problemáticas desde la genuina experiencia de los judíos hispanos, tal como esta se refleja en la filosofía e historiografía de Abraham ibn Daud de Toledo. Ibn Daud es también protagonista, junto a Yehudá Haleví y Moisés Maimónides del estudio del fenómeno de la profecía y de la intelección humana, cuestión de enorme relevancia para el pensamiento judío medieval. La profecía, tal como se aborda en la obra de estos autores, representa el fascinante punto de encuentro entre el intelecto humano y el mundo de la divinidad. Marco en el que la palabra revelada se pone a disposición de las facultades racionales humanas, es un proceso de comunicación único en el que las verdades divinas de naturaleza absoluta son contempladas desde la intrínseca limitación de la mente humana. Merced a su determinación por hacer razonable la experiencia religiosa, los filósofos judíos medievales llevarán la psicología y la epistemología peripatética y neoplatónica, así como sus interpretaciones islámicas, a sus límites conceptuales.

Aunque en la resolución del profetismo los fundamentos de la religión revelada son explicitados con el lenguaje de la racionalidad humana, en un proceso que es comunicación "descendente", sin embargo,

podemos contemplar cómo el judaísmo desarrolla, desde finales de la Antigüedad, un fascinante proceso de comunicación místico con los secretos escondidos de Dios en sentido "ascendente" y que eclosiona con fuerza en el Medioevo. La religión judía, obligada a replantearse su propia esencia en conflicto con la crudeza de los acontecimientos históricos, albergará en su seno formas de experiencia religiosa alternativas al simple devenir del tiempo. En el tercer capítulo del libro, el autor aborda, a partir de la lectura del texto por excelencia de la mística cabalística castellana, el *Zohar* o *Libro del esplendor*, la comprensión de ese nexo que estos hombres medievales creen hallar entre la eternidad desplegada en lo humano y la temporalidad que aspira a lo eterno.

Sin embargo, En el extremo opuesto al polo de las experiencias místicas se halla la búsqueda de una comprensión racional del universo y el hombre creados por Dios. En este sentido, no es exagerado caracterizar la filosofía judía postmaimonideana medieval como un diálogo permanente con la *Guía de perplejos*. El profesor Fernández López se aproxima a este tópico desde el estudio de este fenómeno en uno de los pensadores judíos más originales y profundos de todos los tiempos, Ḥasdai Crescas. La relectura neoplatónica de la cosmología medieval peripatética, la consideración de la cábala como una fuente autorizada de conocimiento religioso, el rechazo a la hermenéutica maimonideana y a la exégesis de las Escrituras que la acompaña, son algunos de los rasgos distintivos de su pensamiento. Una filosofía antifilosófica donde convergen, junto a la concepción judía tradicional, las diversas fuentes aristotélicas, el pensamiento neoplatónico del apóstata Abner de Burgos o la nueva ciencia desarrollada en el siglo XIV por hombres como Jean Buridán y Nicolás Oresme.

A pesar de estas virtudes y más allá del antimaimonedismo de Crescas, es indiscutible que el *Moré Nebujim*, la *Guía de perplejos* ha sido la referencia fundamental en el judaísmo de todo intento posterior de racionalizar la experiencia creyente, un vehículo privilegiado desde el que confrontar la cosmovisión teológica de la Biblia hebrea con el corpus de conocimientos ético-científicos del mundo clásico griego, así como una poderosa herramienta de apoyo a la exégesis de la Torá. En el capítulo quinto, *Estudios* aborda cómo, gracias a la traducción de la *Guía de perplejos* a una lengua vulgar por primera vez en tierras castellanas, la influencia nunca perdida del pensamiento maimonidiano en la Península adquirirá un vigor renovado en un siglo, el XV, en el que, paradójicamente, culmina de forma trágica la historia del judaísmo hispano.

Los capítulos de este libro reflejan, pues, algunos de los hitos de una aventura intelectual y espiritual apasionante, cuyo protagonista es el judaísmo medieval sefardí, una historia donde convergen la experiencia religiosa y la aventura intelectual, la fe y la racionalidad humanas en tiempos de exilio, el pasado y el destino de un pueblo vinculado estrechamente a unas promesas. El destino del judaísmo hispano sufrirá las vicisitudes de una historia general judía traspasada por la intemperancia. A merced de su corriente salvaje, el siglo XV será el tiempo del ocaso definitivo de la vida judía en la Península. Los caminos de un exilio siempre renovado llevarán al judaísmo ibérico a Ámsterdam, al Norte de África, a Italia, los Balcanes y Palestina, y en todos estos lugares, pensadores como León Hebreo, Baruch Spinoza o Isaac Luria reflejarán, de forma tardía, la riqueza y el vigor intelectual de Sefard.

David Soto Carrasco
(Universidad de Murcia)

ORTEGA Y GASSET, J. (2021). *La idea de principio en Leibniz y la evolución de la teoría deductiva*. Madrid: Consejo Superior de Investigaciones Científicas y Fundación Ortega y Gasset-Gregorio Marañón.

La Editorial CSIC y la Fundación José Ortega y Gasset-Gregorio Marañón publicaron en 2021 la segunda edición ampliada, a cargo de Javier Echevarría, de *La idea de principio en Leibniz y la evolución de la teoría deductiva: del optimismo en Leibniz*, una versión que incluye estudios introductorios de Jaime Salas, Concha Roldán y Javier Echevarría.

Esta publicación recoge todo lo prácticamente escrito por Ortega y Gasset sobre Leibniz, pues, además del libro *La idea de principio en Leibniz y la evolución de la teoría deductiva* (1958) y la conferencia *Del optimismo en Leibniz* (San Sebastián, 1947), suma 587 escritos y notas de trabajo que el maestro de la Escuela de Madrid redactó sobre el filósofo alemán, y que hasta ahora habían permanecido inéditas en el Archivo Ortega y Gasset. La intención de incorporar estas notas responde a que muchas de ellas fueron utilizadas por Ortega para la redacción de *La idea de principio en Leibniz*, mientras que otras constituyen borradores que, más tarde, tenía intención de utilizar en los dos tomos que finalmente no fueron publicados, sobre el principio de razón suficiente y el principio de lo mejor. Con esta edición se construyen puentes dialógicos entre dos pensadores y sus tiempos, que coinciden en la representación de unas figuras sobresalientes de la filosofía de su tiempo. La publicación de las 587 notas inéditas simboliza un obsequio para conocer, como Javier Echevarría señala, el modo de pensar y trabajar de Ortega, con el propósito de vislumbrar algunos de los elementos que podrían haber quedado plasmados más tarde. Así pues, este trabajo contribuye al conocimiento sobre el discurrir del modo de pensar orteguiano y su interés en la obra de Leibniz.

A pesar de que esta edición recoge fielmente lo ya publicado en las *Obras Completas de José Ortega y Gasset*, en lo referente a *La idea de principio en Leibniz y la evolución de la teoría deductiva* y *Del Optimismo en Leibniz*, es preciso subrayar que el valor añadido se encuentra en los tres estudios introductorios. En la presente edición han participado importantes especialistas en la obra de Ortega y Leibniz, como son Jaime de Salas (UCM, director del Centro de EStudios Orteguianos), Concha Roldán (Instituto de Filosofía del CSIC, IFS-CSIC) y Javier Echeverría (Jakiunde, Academia de las Ciencias, de las Artes y de las Letras del País Vasco). Como puede comprobar el lector, la Editorial CSIC y la Fundación Ortega-Marañón brindan un trabajo caracterizado por el cuidado y el rigor. Estos estudios, junto a los 587 escritos de Ortega sobre Leibniz, simbolizan una ofrenda para todo lector que se disponga a enriquecer su conocimiento sobre el filósofo español. Sin duda, este trabajo contribuye a ese enriquecimiento y lo hace de la mejor manera posible, homenajeando con tres ensayos que constituyen un encuentro cordial con uno de los filósofos españoles más importantes de nuestra historia.

El primer ensayo, titulado *Ortega en 1947* (pp. 23-47) y escrito por Jaime de Salas, examina las implicaciones de *La idea de Principio en Leibniz* desde una perspectiva de razón histórica e historia de la filoso-

fía del maestro español. De Salas argumenta cómo este texto simboliza la culminación de una obra que ha avanzado con seriedad, se ha ajustado a los problemas contextuales de su tiempo y a diversos elementos teóricos de la tradición. Es fundamental entender que *La idea de Principio en Leibniz* es el resultado de un periodo de madurez, donde el filósofo completa las reflexiones ensayísticas con análisis más sistemáticos y complejos. Es importante reconocer el valor del ejercicio llevado a cabo por Jaime de Salas, que insiste en el carácter histórico de la reflexión orteguiana, haciendo hincapié en algunos de los principales elementos de su obra, que conecta mediante el tejido de un hilo conductor que promueve una visión del texto como una escritura de madurez y consolidación intelectual. Asimismo, es preciso estimar la relación que de Salas establece entre Ortega y otras figuras destacadas de la filosofía como Heidegger, Husserl, Descartes o Aristóteles, con la finalidad de enriquecer la perspectiva histórica que está presente en el pensamiento orteguiano y que es posible percibir desde múltiples aristas.

El segundo ensayo que introduce esta interesante edición ampliada ha sido elaborado por Concha Roldan que, de un modo admirable, lleva a cabo un recorrido intelectual por la historia de Leibniz en el pensamiento de Ortega. Igualmente, sintetiza la recepción de Leibniz en España y el diálogo emprendido por nuestro filósofo con el resto de autores que hay que considerar para entender las implicaciones de su estancia en Alemania: Cassirer, Dilthey, Hartmann, Heidegger o Husserl, entre otros. Recogiendo el testigo de Javier Echeverría, Concha Roldan afirma que el legado leibniziano siempre ha acompañado a Ortega durante toda su obra, hasta el punto de sentirse identificado con el filósofo alemán. Por esta razón, no es de extrañar que Ortega recurra a las tesis del filósofo alemán en numerosas ocasiones. También es interesante la mención a "las enseñanzas subliminales" de Ortega que lleva a cabo Concha Roldan, cuando pone en valor la capacidad del maestro para sortear la censura y los límites impuestos por el nacionalcatolicismo. Finalmente, subraya la influencia de Leibniz en Ortega a través del carácter histórico que el español le confiere al hombre, una idea que constituye uno de los pilares fundamentales de textos como *Historia como sistema* (1938), *La idea de principio en Leibniz* (1947) o *Meditación sobre Europa* (1949). La reflexión de Roldan, que pone de relieve un excelente conocimiento sobre la materia, contribuye a enriquecer el conocimiento sobre el acompañamiento que supuso Leibniz para Ortega (p. 51).

Por último, Javier Echeverría, que puede presumir de ser un gran conocedor de la filosofía orteguiana, nos regala el último ensayo introductorio, titulado *Encuentros de Ortega con Leibniz* (pp. 63-101). Echeverría considera que *La idea de principio en Leibniz* es una gran obra filosófica que demuestra la talla de Ortega y Gasset como pensador. Como se anticipa en el título de este estudio, se abordan los momentos de la vida de Ortega en los que Leibniz se convirtió en una inspiración o, más bien, en su "circunstancia intelectual" (p. 64), en palabras del autor. En el ensayo se profundiza en la relación entre el filósofo español y el alemán, remitiendo al valor que poseen las notas de trabajo de la presente edición. Con la referencia a estas notas, Javier Echeverría argumenta el propósito de incorporarlas a la presente edición, haciendo especial énfasis en la necesidad de conocer el modo de escribir y, por tanto, de pensar de Ortega. De esta forma, asegura que es posible fortalecer el conocimiento de la relación del filósofo español con Leibniz. En mi opinión, una

de las aportaciones más importantes de este ensayo es la claridad con la que son expuestos los momentos del contexto intelectual que configuran el pensamiento de Ortega. Echeverría traza una línea histórica en la que es posible percibir la evolución del proyecto orteguiano, que culmina con la conferencia *Del optimismo en Leibniz* (1947) y la publicación de *La idea de principio en Leibniz y la evolución de la teoría deductiva* (1958).

Estos estudios introductorios integran una edición que ayuda a señalar la actualidad del pensamiento orteguiano y a reconocer la brillantez de su obra. Asimismo, amplían las fronteras del conocimiento de este maestro más allá del género ensayístico, ubicándolo en el espacio de una escritura seria y sistemática. Sin duda, este trabajo no debe pasar desapercibido en posteriores investigaciones que tengan por objeto la presencia de Leibniz en el pensamiento de Ortega y Gasset. Finalmente, animo al lector a encontrarse con esta ilusionante edición, a hacerlo con afecto, poniendo en valor los tres estudios introductorios que facilitan una lectura más sosegada y reflexiva, y a reconocer el esfuerzo que supone el ofrecimiento de los manuscritos del filósofo español relativos a Leibniz.

Antonio Luis Terrones Rodríguez
(Universitat de Valencia / Instituto de Filosofía-CSIC)

COORS, M. (ed.) (2022). *Moralische Dimensionen der Verletzlichkeit des Menschen*. Berlín-Brandenburgo: De Gruyter.[1] [*Dimensiones morales de la vulnerabilidad del ser humano]*.

La reflexión sobre la vulnerabilidad en el terreno de la Ética y la Filosofía política ha ganado relevancia durante las últimas décadas. La crisis del COVID-19 ha reforzado su actualidad, pues ha constituido una experiencia de exposición compartida ante un virus desconocido. No obstante, el debate trasciende la situación pandémica. Así lo demuestra este recién publicado volumen, fruto de seis reuniones organizadas por el proyecto de investigación financiado por el Centro de Ética de Salud —*Zentrum für Gesundheitsethik*— de Hannover[2] y el Instituto de Ética social[3] —*Institut für Sozialethik*— de la Universidad de Zúrich entre 2018 y 2020. Estamos, así, ante una propuesta de reflexión de gran interés sobre uno de los conceptos más relevantes en las discusiones éticas de las últimas décadas. Estando escrito en alemán, que la barrera lingüística impidiera al ámbito hispano acceder a las líneas de investigación de este

1 Este trabajo ha sido posible gracias a las ayudas predoctorales para investigación y docencia "Programa Severo Ochoa" (BP20-147) del Principado de Asturias.

2 Para más información, consúltese la página web del centro: https://www.zfg-hannover.de

3 Para más información, consúltese la página web del Instituto: https://www.asae.uzh.ch/de.html?gclid=Cj0KCQiAj4ecBhD3ARIsAM4Q_jG0YTL7Z-SE8-heiWK5eRCzwldoPAcrMMAafkccglkpI1-eDzz-c5waAo6sEALw_wcB

volumen conllevaría una verdadera pérdida. Precisamente por ello nace esta reseña.

El libro es introducido por Michael Coors (pp. 1-23), quien abre el debate sobre la vulnerabilidad y discute cómo dicho concepto se modula de acuerdo a las coordenadas sociales, económicas y simbólicas en que se contextualice: la edad, la clase, la etnia o el género son factores de riesgo que evidencian cómo la vulnerabilidad se vincula con "cuestiones de justicia" (p. 2). Así, el prólogo plantea la serie de preguntas que recorrerán el resto de propuestas. Y es que la vulnerabilidad demuestra ser una idea-eje cuando se trata de reflexionar sobre aspectos éticos, políticos, legales y, también, existenciales. La cuestión es que la propia definición de la vulnerabilidad exige un debate. Al respecto, Coors y el resto de autores partirán de una premisa común: no puede ser reducida a una suerte de incapacidad individual para defenderse, pues esto implicaría olvidar las dimensiones sociales que la definen. No en vano, la vulnerabilidad comienza a tomar especial importancia con el surgimiento de las éticas del cuidado de raigambre feminista o comunitarista, con autores como Martha Nussbaum o Alasdair MacIntyre; de acuerdo con ellos, el ser humano se define por su carácter inextricablemente vulnerable desde el nacimiento, y de dicho estatus surgen, precisamente, las obligaciones morales hacia los otros.

En primer lugar, Burkhard Liebsch (pp. 27-55) presenta por escrito una conferencia donde analiza la vulnerabilidad como una constante existencial: el ser humano *es* vulnerable y durante su vida adquiere *otras* vulnerabilidades. Será esta una base común en el resto del libro. Además, nos habla de una vulnerabilidad en sentido extremo: aquella que pone al límite las condiciones físico-psíquicas y que se topa con las fronteras del lenguaje, por decirlo con Wittgenstein. Esa vulnerabilidad no verbalizable es el telón de fondo del resto de vulnerabilidades. Qué sea en verdad es, pues, algo que solo podemos comprender por experiencias tentativas. Pero, al tiempo, es la obligación de toda comunidad tratar de comprender ese horizonte "asintótico", por así decirlo, de vulnerabilidad en el que irremediablemente nos movemos; ese confín incomunicable pero, a la vez, constitutivo del ser humano. Más aún, pensando con Hegel, se da la circunstancia de que sufrir y hacer sufrir es inevitable, por cuanto el ser humano es inextricablemente político, y de que tal es la única manera de comprender aproximadamente qué es la vulnerabilidad para poder generar límites morales y legales. Y es que aunque en el "orden de los hechos" la vulnerabilidad sea la condición del daño, a escala cognitiva el orden es el inverso. Por lo tanto, la vulnerabilidad, en condiciones adecuadas, nos permite tanto autoconocernos como comprender la existencia del daño ajeno. Más aún en una sociedad secularizada donde el imaginario del infierno, que no dejaba de constituir una manera de acercarse simbólicamente a la extrema vulnerabilidad, ha perdido su sentido; se precisan otras vías de pensamiento para aprehender esta realidad tan escurridiza como, a la vez, definitoria de los seres humanos.

En una línea similar participa Rebekka A. Klein (pp. 57-84), defendiendo la necesidad de una "antropología encarnada": la vulnerabilidad no se reduce a situaciones de dolor concretas, sino que es un espectro que siempre está en el "fondo antropológico". En otras palabras, no es un atributo o complemento del sujeto, sino nuestra estructura existencial. Razonar exclusivamente en términos de vulnerabilidad específica conlleva estigmatizar a los grupos sociales minoritarios; es decir, olvidar el sentido antropológico de la vulnerabilidad deriva en políticas

paternalistas. Así, en vez de concebirla, como suele ser la tendencia, como un concepto negativo o privativo, debe ser tomada, en su carácter ambivalente, como un rasgo constitutivo del propio yo. En definitiva, la vulnerabilidad no siempre consiste en un daño, sino que puede abrir nuevas sendas de desarrollo individual y colectivo, esto es, formas de empoderamiento alternativas a la idea tradicional del sujeto abstracto y auto-soberano, y verdaderamente sensibles a la violencia hacia el Otro en nuestra radical imperfección.

En el texto del propio Coors (pp. 85-103) confirmamos que el debate sobre la vulnerabilidad exige superar la definición de estela kantiana del sujeto como un ser autónomo, dotado de capacidad de decisión libre y racional, para reconocer su identidad relacional. La comprensión formalista del individuo olvida el carácter encarnado de la moral; el propio término "vulnerabilidad" incluye en su sentido etimológico (*vulnus*, "herida") dicha dimensión. Así, se exige ir más allá de una ética del deber donde el eje principal sea la autodeterminación, para ponerla en tensión con la vulnerabilidad y su dimensión corporal y psicosocial. Por este motivo, Coors dedica parte del capítulo a repasar las éticas de raigambre liberal y comunitaria. Mientras que la autonomía suele asociarse con la dimensión "racional", la vulnerabilidad se relaciona de forma más habitual con nuestra parte "física". Sin embargo, si somos coherentes con la irrenunciable crítica al dualismo realizada por la filosofía contemporánea, se evidencia que la vulnerabilidad permitirá no solo considerar la cuestión del cuidado, sino también valorar y proteger la libertad y autodeterminación ajenas.

A continuación, Noelia Bueno Gómez (pp. 105-126) se centra en la cuestión del sufrimiento desde una perspectiva biopolítica. La necesidad de este enfoque se evidencia en cuanto se considera que son los mecanismos del biopoder los encargados de gestionar el sufrimiento y los procesos de vulnerabilidad en una comunidad. Bueno defiende, frente a otras "sociodiceas" -p.ej., el catolicismo-, que el sufrimiento es una experiencia negativa; sin afirmar que deba ser abolido, debe ser "tomado en serio". La *polis* debe ser diseñada de tal forma que no produzca nuevas situaciones de vulnerabilidad derivadas de la instrumentalización del sufrimiento para fines no vinculados con el acabamiento de dicho dolor. Constatar esto permite a Bueno establecer unas líneas críticas. En primer lugar, asociar un grupo con una vulnerabilidad puede conducir al paternalismo y la estigmatización, además del olvido de otros grupos. Se evidencia, además, la necesidad de superar la comprensión del sufrimiento en términos exclusivamente médicos. La cuestión sería hallar mecanismos de biopoder que no constituyan una manipulación instrumental de la vida.

La segunda parte del volumen tiene una intencionalidad más cercana a la ética aplicada y la bioética y se centra en formas de "vulnerabilidad contingente". En este sentido, la contribución de Tobias Eichinger (pp. 127-142) funciona a modo de bisagra, pues su campo de aplicación es la medicina, pero aún de forma amplia y teórica. El concepto de salud debe ser relacionado específicamente con el de vulnerabilidad para comprender la enfermedad, el daño y la degeneración vital de los cuerpos. Al respecto, analiza la "biomoralidad" o "sanitarismo": esa exigencia de ser feliz y saludable propia de la sociedad contemporánea, que fuerza al lenguaje médico a contener un mensaje moralizante. Es cierto que ha tenido como cara positiva el hecho de trascender una visión de la salud negativa, estrictamente física y centrada en *evitar* la enfermedad, para pasar a incluir el bien-

estar psicológico y social. Pero la falacia naturalista que se incluye detrás de muchos mensajes médicos debe ser criticada, junto con las formas de mercantilización de la salud traídas por el capitalismo.

Prosigue Claudia Bozzaro (pp. 145-163) centrándose en la idea de dolor. Primero, estudia el agudo pero transitorio, que le permite defender la vulnerabilidad como una condición humana genérica: siempre estamos expuestos al riesgo, e incluso es necesario -e inevitable- para comprender el carácter corporal de la propia identidad. Por lo tanto, este primer tipo de dolor tiene un carácter ambivalente, pues aun siendo una experiencia evidentemente negativa, es temporal y la reacción que genera es necesaria para la supervivencia. A continuación, se detiene en el dolor crónico, esto es, el permanente y, a menudo, no curable. En este caso, la función de la supervivencia no existe y padecerlo produce una "cascada de vulnerabilidad": en muchos casos, la enfermedad pasa a ser el eje del propio proyecto vital, produciendo a menudo otros males de tipo psicológico o social. Además, los dolores que provoca a su alrededor no siempre son de una localidad clara y, por lo tanto, son difíciles de aliviar. Con este análisis en dos partes, Bozzaro propone una comprensión gradual de la vulnerabilidad.

Andrea Dörries (pp. 165-177) se detiene en el caso de los bebés prematuros y la atención sanitaria que precisan. Los bebés prematuros son claramente vulnerables -carecen de autonomía para comer o incluso respirar, apenas se pueden mover, etc.- pero, al tiempo, muestran una capacidad de resiliencia que permea el imaginario social. El cuidado de evitar la exposición de estos bebés es también reflejo de la serie de conductas médicas y políticas que suelen aplicarse para prevenir la aparición de vulnerabilidades específicas. Al respecto, ha habido un cambio en la concepción de los bebés prematuros desde los años 70, gracias a las tecnologías de respiración artificial y otras novedades médicas: se ha pasado de una evasión del cuidado a un cuidado "a fondo" que tiene en cuenta el contexto familiar. Sin negar lo positivo del cambio, surge como pregunta en qué condiciones iniciar los cuidados paliativos, que ha llevado a distintos criterios en diferentes legislaciones médicas. Más aún, las propuestas de una menor incubación suelen ser fuertemente criticadas, a pesar de estar demostrado que el contacto con la piel es beneficioso para el desarrollo del bebé. Así, la comprensión de su vulnerabilidad está condicionada por el conocimiento de la época, que determina la visión simbólica de esta fase de la vida.

La cuestión de la disposición genética humana corre a cargo de Henriette Krug (pp. 179-203). Ser parte de una familia con una enfermedad hereditaria genera un sentimiento de incertidumbre radical en sus miembros que les fuerza a experimentar la vulnerabilidad incluso cuando no han llegado a contraer dicha enfermedad, especialmente en los momentos previos a los resultados de las pruebas médicas. Krug pone el ejemplo de la enfermedad de Huntington a la hora de estudiar esa experiencia de poseer una vulnerabilidad "genética". Las familias que la padecen viven en un clima de conciencia de vulnerabilidad intensificada, pues los miembros que aún no la han contraído son testigos de quienes sí y, por lo tanto, de su propio futuro. Los resultados de la prueba modifican el autoconcepto y relaciones con los demás y, en definitiva, el propio proyecto de vida, pudiendo conducir hacia nuevas enfermedades, por ejemplo, de tipo psicológico. Entramos, de nuevo, en la existencia de una cascada de vulnerabilidad.

El envejecimiento es otra arista habitualmente asociada culturalmente con la vul-

nerabilidad, examinada por Mark Schweda (pp. 205-227). El autor analiza críticamente dicha vinculación, cuestionando si las personas de edad avanzada pueden ser efectivamente consideradas como un grupo social. Se suele relacionar el envejecimiento con la degeneración cognitiva, cuando no guardan una relación paralela necesaria. Como discute el autor, no hay un vínculo directamente evidente entre la vulnerabilidad y la vejez. De hecho, con el aumento de la esperanza de vida, de los derechos de asistencia y la tendencia de la población envejecida a tener una mayor solvencia económica que la juventud, no necesariamente constituye un grupo social vulnerable. Su vulnerabilidad se da por circunstancias sociales más que por el proceso de envejecimiento.

El final de la vida es, sin duda, el momento de vulnerabilidad que, a nivel existencial, más tenemos presente los seres humanos. Seguramente no es casualidad que sea el cierre del libro, a cargo de Christoph Rehmann-Sutter (pp. 229-248). Distingue la muerte violenta, producida por el uso de la fuerza, negligencia o alguna causa externa, de una que "nace de la vida": la muerte inevitable, incausada. La muerte inevitable es, pues, parte de la identidad humana, y no puede ser entendida como una lesión. Sin duda, qué se pueda entender por este tipo de muerte y su necesaria demarcación con lo que se suele denominar "muerte natural" –que no siempre es inevitable– es interpretable, y el autor se detiene en problematizar debidamente el concepto. Oímos ecos epicúreos al leer que la vulnerabilidad no es la muerte per se, sino el sufrimiento que pueda experimentarse en su antesala, y que tal debiera ser el objeto reflexivo de la ética. El deseo de morir pasa, así, a ser el final de sus reflexiones, examinando cómo la desatención de dicho deseo, pero también de las circunstancias que han podido conducir hasta él, sí puede generar vulnerabilidades específicas.

En definitiva, el lector podrá acceder a una cuidadosa selección de análisis y miradas hacia este fenómeno que, lejos de ser definible de una vez y para siempre, se nos aparece, tras la lectura, como una suerte de red: tal malla es la vulnerabilidad "antropológica", y recorre todo cuerpo individual y social, recordándonos que somos irremediablemente finitos. Al tiempo, toda red tiene unos nudos que la distinguen de otras redes; y así, cada ser humano, *siempre* vulnerable, también se *hace* vulnerable por otras razones y en distintos momentos del curso vital. Los autores han seguido varias de sus bifurcaciones. Queda al lector reconectarlas con las propias.

Isabel Argüelles Rozada
(Universidad de Oviedo)

MBEMBE, A. (2022). *Brutalismo.* Traducción de Núria Petit. Barcelona: Paidós.

En su último libro, titulado *Brutalismo*, publicado en Francia en 2020 y traducido en 2022 al castellano, el filósofo camerunés Achille Mbembe describe unas estructuras heredadas de la economía colonialista comparables en durabilidad y fortaleza al mencionado estilo arquitectónico. Asimismo, examina la historia del colonialismo europeo

en África y la diáspora africana provocada por la esclavitud, el presente de estas sociedades y las condiciones que deben darse para que en el futuro se apoyen en cimientos más respetuosos y prósperos.

Mbembe es un filósofo asentado en la teoría crítica del capitalismo, de modo que en sus análisis no cree necesario justificar el beneficio económico en todo momento para todas las partes: las causas justas no tienen la obligación de ser rentables. En una sociedad como la actual en la que predominan los argumentos económicos, ésta podría ser a simple vista una de las debilidades de este libro en particular y de la obra de Mbembe en general: la renuncia a supeditar la moralidad a propuestas concretas. En sus páginas no encontraremos un análisis de la inmigración africana en relación a poblaciones como la europea con problemas de crecimiento; ni se pregunta quién pagará las pensiones en un alegato a favor de unas poblaciones jóvenes inmigrantes que utilizan los recursos públicos en un porcentaje mucho menor que las envejecidas poblaciones locales (Banerjee y Duflo, 2019, 10-23). Este tipo de argumentos supondría una absolución parcial a un proceso que considera injusto en su totalidad.

En consecuencia, su intención es extender una línea temporal entre las democracias liberales y las sociedades colonialistas, enfrentándolas a un espejo que refleje la forma en la que regulan el flujo migratorio mediante perversas medidas fronterizas (Mbembe, 2022, p. 125). Debido a que algunas son demasiado reveladoras incluso para la acomodaticia hipocresía occidental, las externalizan como forma de exculpación frente a sus votantes. No resulta extraño, por tanto, que las tragedias más mediáticas tengan lugar en un medio natural— el mar—en el que esta externalización resulta más complicada.

En opinión de Mbembe, las democracias liberales deben enfrentarse a la contradicción de que mientras en otras épocas desplazaban poblaciones completas a la fuerza para poblar sus colonias «por temor a la despoblación, es decir, a las condiciones que hacen posible que la especie humana se extinga» (Mbembe, 2022, p. 111) o como método de producción preindustrial al promover la esclavitud, en la actualidad empleen ingentes cantidades de recursos en evitar su entrada.

Los antiguos poderes coloniales se defenderán mostrando orgullosos la socialdemocracia que intentan exportar al mundo. De nuevo la asimetría y la obsesión con exportar—da igual que sean conceptos, mercancías o ideas—mientras se frena la libre circulación de las personas, sin importar que el artículo 13 de la Declaración Universal de los Derechos Humanos de 1948 recoja el derecho a la libre circulación y a elegir el lugar de residencia dentro del propio estado.

Si bien Mbembe utiliza este artículo como prueba del incumplimiento de las promesas de las socialdemocracias, tal vez sería más apropiado verlo como un ladrillo *brutalista* debido a que deja en manos de los estados la regulación de la inmigración al no contemplar el derecho a emigrar entre fronteras. Resulta paradójico que una declaración universal sea tan respetuosa con las jerarquías estatales, de no ser porque el objetivo de los redactores del artículo 13 fue «evitar deportaciones masivas de población, como las ordenadas durante el estalinismo y el nazismo» (Arcos Ramírez, 2020, p. 4). En resumen, el objetivo de dicha declaración fue solucionar los numerosos y muy diversos problemas internos europeos derivados de las guerras y no tanto los relacionados con el colonialismo. El *Brutalismo* colonialista también se manifiesta en los problemas que no se consideran prioritarios.

¿Cómo funciona esta estructura? En primer lugar, ejerciendo «si es preciso por la fuerza» (Mbembe, 2022, p. 11), una dominación mediante unos sistemas e instituciones que demuelen estructuras nativas y construyen con l*adrillos elementales* (Mbembe, 2022, p. 11), creando «a escala planetaria reservas de oscuridad» (Mbembe, 2022, p. 11). En segundo, mediante una firmeza que contrastará con la vulnerabilidad y precariedad de poblaciones enteras que serán lanzadas contra unas fronteras que servirán a la vez de poderosa atracción y repulsión. Por consiguiente, la frontera será una promesa de vida al compararla con las condiciones que los inmigrantes dejan en sus países de origen, «cuyos medios de supervivencia se han destruido» (Mbembe, 2022, p. 58) pero, a la vez, los tratos vejatorios y dificultades del camino serán un poderoso anticipo de la exclusión que les espera en su destino (Mbembe, 2022, p. 11). Un destino que podría no ir más allá de la propia frontera y su inacabable área de influencia, pues «la fronterización ha transformado estos espacios en lugares infranqueables» (Mbembe, 2022, p. 58).

Cabe mencionar que Mbembe no limita a África su crítica al neoliberalismo, sino que también lo culpa de «la combustión del mundo…y el agotamiento vertiginoso de los recursos naturales…que sostienen la infraestructura material de nuestra existencia» (Mbembe, 2022, p. 11). La humanidad ha vivido sin poner límite a su existencia sobre el planeta, asumiendo que los recursos naturales eran algo a explotar a través de su ingenio, pero según Mbembe «estamos de lleno en la edad de combustión del mundo» (Mbembe, 2022, p. 21). Debemos cambiar y moldear nuestro consumo, tomando conciencia de que habitamos un planeta con recursos limitados. Mientras no aceptemos esta limitación, ninguna explotación será verdaderamente eficiente.

Esta toma de conciencia no está exenta de peligros, como los deseos excluyentes de naciones autóctonas—tan presentes en una extrema derecha en auge—y que provocan una «reactivación a escala planetaria del deseo de endogamia y de las prácticas de selección y triaje» (Mbembe, 2022, p. 22). En consecuencia, las fronteras serán "aparatos de captura, de inmovilización, de alejamiento de poblaciones consideradas indeseables" (Mbembe, 2022, p. 111).

Del mismo modo que el tráfico de esclavos cambió la fisonomía del mundo, Mbembe nos recuerda que la diáspora africana no ha dejado de producirse debido a las condiciones de vida de unos países de origen que sufren los efectos de primar el beneficio empresarial por encima de las vidas de las personas con una «desregulación de las transacciones financieras [y la] sumisión de los servicios públicos [a] las condiciones de rentabilidad del sector privado» (Mbembe, 2022, p. 111).

Paradójicamente, la misma obsesión productiva que creó las condiciones para la esclavitud, se enfrenta ahora a que esa mano de obra ya no es necesaria en unas sociedades mecanizadas, de modo que estas poblaciones son prescindibles, realizándose una «distinción entre personas humanas solventes e insolventes» (Mbembe, 2022, p. 111). Acusa al capitalismo de ni siquiera mantener rasgos de caridad de sistemas anteriores que asociaban estas figuras sobrantes con «figuras crística…objeto de cuidados caritativos» (Mbembe, 2022, p. 114). La creación de instituciones asistenciales y de cambios legales criminalizando el vagabundeo contribuyó a la penalización del derecho al libre movimiento. En las épocas imperiales la vocación universalista de la iglesia convertía a los mendigos errantes

en un reflejo de la divinidad—de ahí la creación de caminos de peregrinación—dotando de una dimensión moral a la sociedad de la que el *Brutalismo* capitalista carece. En consecuencia, mientras la expansión fue una prioridad, bien de una religión o de una nación con afán colonialista, la libertad de movimiento e incluso los cuidados básicos fueron respetados, mientras que al perderse estas motivaciones «el tratamiento de los cuerpos migrantes asimilados a cuerpos virulentos o a desechos humanos irá adoptando progresivamente la apariencia de profilaxis social (Mbembe, 2022, p. 115)». En otras palabras, el racismo disfrazado de higiene, la discriminación encubierta por un supuesto avance social encubriendo que la sociedad receptora considera «fuentes de potenciales molestias» (Mbembe, 2022, p. 125), a quienes en otro tiempo categorizó como recursos aprovechables.

Para Mbembe el capitalismo y el colonialismo son dos cabezas del mismo monstruo. En unos capítulos con lenguaje sexual—con títulos como *Virilismo o Phallos*— (Mbembe, 2022, pp.87-106) en los que se sirve tanto del imaginario africano como el demoniaco que el colonialismo quiso asociar a las poblaciones autóctonas; en un intento de suspender en las colonias la moral tradicional y puritana de los países de origen—donde la iglesia regulaba la relación íntima entre las personas—convirtiéndolas en lugares en los que llevar a cabo todo tipo de perversiones. Describe un colonialismo patriarcal y violador, un sistema en el que el falo no sólo es «el emblema y la insignia del poder» (Mbembe 2022, p. 97), sino que es el poder en sí. Una *falocracia* cuyo volumen se dilatará y contraerá en el ejercicio de la opresión y que se sentirá amenazada por la virilidad de los esclavos. De este modo, la exaltación de dicha virilidad será en el régimen de la plantación y del gobierno colonial una forma de deshumanizar y desvirilizar, aplicando castigos en los que el hombre negro era separado de su miembro precisamente como forma de demostrar que era «ante todo un miembro» (Mbembe, 2022, p. 98).

En cuanto a la organización de estas *falocracias,* uno de sus pilares básicos era la *deuda familiar;* «la deuda de los hijos con respecto a los padres y la idea de complementariedad en la desigualdad entre hombres y mujeres» (Mbembe, 2022, p. 101); una deuda común contraída con las generaciones anteriores y que, para ser pagada de la mejor forma posible, tenía que ser distribuida según parámetros tradicionales. Para Mbembe estos conceptos ya no están vigentes y se ha producido una desmasculinización en las sociedades africanas la cual, paradójicamente, ha facilitado la masculinización de las clases dirigentes convirtiendo el dominio «en el privilegio de unos pocos» (Mbembe, 2022, p. 101). Esto explicaría que en unas sociedades con ideas cada vez más igualitarias, las esferas de poder sigan comportándose con violencia *falocrática* y ejerciendo, respecto a sus subordinados, de «padre dentro de la unidad familiar". (Mbembe, 2022, p. 101).

Al respecto de futuras reparaciones de los bienes museísticos expoliados por las potencias coloniales, Mbembe recuerda que no se puede vivir con el resentimiento y que las sociedades africanas aprenderán «a vivir con esa pérdida» (Mbembe, 2022, p. 303), pero que las naciones europeas deben a su vez aprender a convivir con el reconocimiento de cómo fueron obtenidos estos objetos, erradicando cualquier condescendencia que agrave el insulto cuestionando la capacidad de los museos africanos de custodiar su propio arte. Al hacerlo, se está justificando el expolio, como si la ciencia

museística occidental hubiera salvado estos objetos de la desaparición.

Éste es el tipo de expolio no sólo material, sino espiritual al que se refiere Mbembe, considerando fundamental tener *capacidad de verdad*, sin la cual no es posible la verdadera restitución (Mbembe, 2022, p. 189). Mientras esta verdadera intención de restitución no se produzca, dichas reparaciones serán superficiales y poco fiables e incluso pueden servir el propósito opuesto al deseado si Europa concluye que esta acción priva a las poblaciones africanas «del derecho a recordarle la verdad» (Mbembe, 2022, p. 190).

En conclusión, las democracias liberales suponen para Mbembe una conveniente e interesada absolución que permite olvidar los abusos del colonialismo europeo en África y de la esclavitud en Estados Unidos. Más allá de reparar este expolio devolviendo piezas de museo, es necesario ser conscientes de aquella realidad y de cómo aún hoy se está impidiendo el desarrollo del continente africano con los controles de los flujos migratorios que obligan a muchos de sus habitantes a vagar por las fronteras, atraídos por el ideal de una vida mejor, pero repelidos por la realidad egoísta de un continente que se niega a aceptar sus responsabilidades históricas. Sólo a partir de este reconocimiento, argumenta, se podrán derribar las estructuras *brutalistas* que permitieron y siguen permitiendo dicho expolio de recursos económicos, espirituales y naturales.

Referencias

Arcos Ramírez, Federico (2020). ¿Existe un derecho humano a inmigrar? Una crítica del argumento de continuidad lógica. *Doxa, Cuadernos de Filosofía del Derecho*, 43, 285-312. Recuperado de https://rua.ua.es/dspace/bitstream/10045/106924/1/Doxa_2020_43_11.pdf

Banerjee, Abhijit V. y Duflo, Esther (2019). *Good Economics for Hard Times*. Nueva York: Public Affairs

Mbembe, Achille (2022). *Brutalismo*. Barcelona: Paidós.

David Alexis Ferrá Vallés
(Universidad de les Illes Balears)

GROYS, B. (2022). *Filosofía del cuidado*. Buenos Aires: Caja Negra.

En el mundo de la inmediatez, en el que la preocupación solo se ocupa del presente ¿qué significa cuidar? En una sociedad burocratizada e institucionalizada ¿cómo puede el hombre cuidar de si?, ¿se posee la libertad para cuidarnos o la estamos depositando en el otro? Boris Groys, en apenas 134 páginas es capaz de acercarnos al cuidado en su máxima expresión: devuelve el poder de cuidar al sujeto. Lo extrae de la perversión del sistema, lo destecnologiza y lo proyecta más allá de la humanización. En una especie de genealogía, el autor hace una crítica al fenómeno del cuidado como propiedad del ser. Más específicamente como un asunto del *Dasein*. Pero para ello, invita

a una lectura de varios autores desde Platón hasta Aleksándr Bogdánov, que han problematizado el cuidado desde la óptica filosófica. El retorno a una crítica del cuidado a partir de las preguntas: ¿qué se cuida?, ¿quién cuida? y ¿cómo cuida? Son apenas los hilos conductores de ese recorrido que intrinca la filosofía del cuidado. Además, Boris, como crítico de arte, se atreve a comparar el escenario del arte: el museo con el lugar en el que nos cuidan: el hospital. Y no precisamente en sentido positivo, sino que alude directamente a la medicalización de la asistencia y a ser tratados como "objetos": el hospital aquí es un observatorio del cuerpo y, por tanto, un dispositivo de control. Y es que, las necesidades del sujeto del cuidado trascienden de lo que puede ofrecernos un centro sanitario, porque el cuidado pertenece al humano por el hecho de serlo.

El libro está compuesto de doce capítulos y uno introductorio. En este primer acercamiento al fenómeno del cuidado, el autor, recupera la afirmación de Foucault: «la salud sustituye a la salvación» para situar el papel de las instituciones del cuidado. En el capítulo introductorio («cuidado y cuidado de si»), el autor hace una interesante reflexión en torno a lo que ha supuesto institucionalizar el cuidado. Define el concepto de *cuerpo simbólico* como una extensión de nuestro cuerpo físico. Es la puerta de entrada a los centros de cuidado. Nuestro cuidado está mediado por estos *cuerpos simbólicos:* pasaporte, DNI, historial médico y otros documentos de identidad. Para Groys, lo anterior simboliza que, en realidad, el cuidado está integrado en un sistema cuyo interés trasciende de nuestra propia salud. Y es en este punto en el que cobra sentido lo que Helmuth Plessner denominaba *sujeto excéntrico* y que rescata Boris, para posicionar el cuidado de sí y relacionarlo con lo que defendía Foucault sobre el papel de las instituciones. Afirma que, aunque la medicina se ha erigido como ciencia, la elección de un tratamiento médico por parte de un paciente supone más un acto de fe irracional. Por la sencilla razón de que nosotros, los que somos cuidados, no poseemos conocimiento médico. Como cuidadores de nosotros mismos asumimos una posición excéntrica, es decir, externa y exenta de ese saber que posee el otro. Además, el hecho de ser un sujeto del cuidado de sí, no me hace disponer acerca de cómo cuidarme, sino que me convierte en un objeto del cuidado: hoy día, paradójicamente, se relaciona cuidar de si con el seguimiento de un tratamiento, con soportar pasivamente todo procedimiento, es decir, solo los que luchan se convertirán en verdaderos cuidadores (héroes). Obviamente esto tiene que ver con los poderes sociales (*das Volk*) y la irrupción del capitalismo. Retornar la participación activa del sujeto del cuidado de si en la toma de decisiones políticas, médicas y administrativas relativas al cuidado de su cuerpo, supondría la ruina de este sistema creado en el seno de la modernidad. De manera singular sus observaciones posrevolucionarios y postindustriales desde Hegel y Nietzsche, muestra como la historia, en su dialéctica y más específicamente en su negación de la negación, el hombre, una vez libre de la hegemonía monárquica y teológica; recurre al Estado. Para que instituya el dispositivo de cuidado. Es entonces cuando Boris, no solo intenta, si no que logra desarticular la perspectiva simbólica del cuidado. Pasando de cuidar cuerpos a cuerpos simbólicos, reivindicando la noción *foucaultiana* del cuidado de sí, dando pruebas prácticas de cómo se pasa de un cuidado a un cuidado de si y viceversa (de un cuidado de si a un cuidado). Este aparente cambio de palabras supone, específicamente, lo que está en juego, según el autor.

La obra hace una diferencia de estas dos nociones y el autor se encarga de entrar y salir, de subir y bajar este concepto de cuidado en los distintos niveles y a partir de varios ejemplos. Niveles en el sentido de situar al Estado y al sujeto; el primero como productor de la subjetividad a partir de la institucionalización del cuidado y el segundo como responsable auténtico de tal cuidado. A modo de un uróboro, el autor le da la vuelta a la noción de cuidado, dejando en una situación un tanto incómoda o nauseabunda al lector. Ya que, según él, el sujeto del cuidado de sí solo puede ser dado a partir de cómo la institución funda en este el modo de cuidar. De tal forma que los últimos capítulos del libro se orientan a poner de relieve cómo es la vida de tal sujeto según su lugar de objeto o sujeto del cuidado (cuidado de sí). El libro por su sentido, comienza y acaba con una oda a la liberta de cuidar de si o lo que él llama el *cuidado revolucionario*: poner en el centro la evidencia y la experiencia del paciente y esto implica inexcusablemente un diálogo entre saberes institucionalizados y cotidianos (*das Volk*).

A partir de aquí, Boris, en los capítulos venideros analiza cómo se posiciona el «yo cuidador» frente y desde lo externo que conforma el mundo en el que vivimos.

Entre las reflexiones del autor nos gustaría destacar algunas que, a nuestro modo de ver, resultan especialmente interesantes.

Pone en tela de juicio la posición de no-conocimiento que se le ha adjudicado a los pacientes. Argumenta que, paradójicamente, solo ellos guardan el saber para el cuidado de sí. Entonces la pregunta aquí es: ¿Por qué buscamos el cuidado fuera? El autor lo tiene claro: el capitalismo y el consumismo nos han hecho ser dependientes de lo material del cuerpo. Obviamente esto es una debilidad, pues nos convierte en sujetos pasivos (objetos del cuidado). A partir de aquí, reflexiona acerca de la manifestación de la libertad, ya que solo así podremos llegar a ser sujetos del cuidado. Y lo hace a través de la dialéctica del amo y del esclavo de Hegel: el sujeto del cuidado de si a diferencia del sujeto del cuidado busca la libertad, la emancipación estatal. En este sentido, tal búsqueda es una liberación. Con su dialéctica y doble negación, el sujeto del cuidado de si no es posible. En tanto que el cuidado de si también es instituido por el disciplinamiento del sistema. El sujeto es, pues, central en la comprensión del cuidado y del cuidado de si, de ahí su pertinencia en tanto cosa o ser-ahí. Pues esta existencia (*Dasein*) es definitoria para el autor.

Para Boris, el cuidador supremo es el pueblo, pero este está preso por la tecnologización o en este caso, la medicalización de la vida cotidiana. La existencia está mediada por la inmediatez, vivimos el aquí y el ahora, únicamente nos preocupamos del presente, la predisposición a vivir un mañana ha desaparecido, en todo caso hemos cambiado la eternidad por un futuro en el que todo lo anterior, lo de otra época, desaparece. ¿Cómo podemos entonces recuperar la tan ansiada autonomía? Poniendo en valor la subjetividad individual o lo que es lo mismo: manifestando nuestra libertad. Para ello el autor, recurre al *Dasein*, el modo de definir al hombre de Heidegger: un ser que se preocupa por la existencia porque es consciente de la amenaza de la muerte. Por tanto, si me preocupo por la existencia, cuido de ella. Por eso, el ser del *Dasein* es el cuidado. Es decir, seremos libres y sujetos del cuidado en el momento en el que manifestemos nuestra voluntad de poder, que reconoce la historia, la mira con ojos del presente y evidencia las diferencias creando algo nuevo (creatividad), pero cuidado, teniendo en cuenta que el *Dasein* solo es, en relación con su contexto y su mundo.

Cuidar no es únicamente conservar nuestros cuerpos, proteger nuestra «vida útil». Cuidar es volver a la esencia de lo humano: al arte, los símbolos y ritos, las tradiciones populares, es volver a reencontrarnos con lo que fuimos, lo que somos y lo que queremos ser. La verdadera revolución del cuidado surge de lo cotidiano, de nosotros, los que cuidamos. Un cambio en los sistemas-Estado, es posible siempre que adoptemos, como remarca el autor, una posición excéntrica que transforme las necesidades de la sociedad y las enfoque a las de las personas que son cuidadas y cuidan. La cultura del paciente como persona que espera esa promesa de felicidad, tiene urgentemente que ser transformada por la de sujeto del cuidado.

El libro es, en realidad, apenas una introducción a la problematización del cuidado, podríamos decir «una introducción a la filosofía del cuidado». Pero una introducción demasiado compleja en tanto que es filosófica y contemporánea. Qué con agudos comentarios invita al lector a replantearse las formas de cosificación del sujeto, a partir del cuidado de sus cuerpos simbólicos, abriendo la crítica a formas de subjetivación contemporáneas. Queda resolver como la salud supone una «Gran infección» y de qué modo el trabajo antepuesto a la labor procura el cuidado revolucionario. Si bien, el libro abre con la cita de Foucault a su arqueología, es de cierto modo una genealogía de cuidado. Una crítica al cuidado moderno, a la modernidad del cuidado.

La obra de arte muere al ser cuidada por la industria, se cosifica en un museo. De igual modo, el ser humano enferma y muere en la medida que es alejado de su cuidador auténtico: el pueblo.

Altamira Camacho, Ramiro
(Universidad Autónoma de Aguascalientes, México. https://orcid.org/0000-0003-3403-6901)

Herrera Justicia, Sonia
(Fundación Index, Granada, España. https://orcid.org/0000-0001-7977-6781)

ZAMORA BONILLA, J. (2022). *En busca del yo. El mito del sujeto y el libre albedrío*. Barcelona: Shackleton Books, 176 pp.

Cuando a finales del siglo XVIII el idealismo de Fichte se derramó sobre el público alemán, hubo unos pocos pícaros que quisieron entender su noción de Yo como una alusión explícita a sí mismo y defendieron la absurda idea de que la *Doctrina de la ciencia* venía a ser una glosa filosófica a su ensortijada vida conyugal, llegando a preguntarse seriamente, en una nota periodística, cuál podría haber sido la reacción de la señora de Fichte contra su cónyuge en represalia por el estorboso papel de No-Yo que éste le habría asignado en su obra.

El libro *En busca del yo. El mito del sujeto y el libre albedrío* de Jesús Zamora Bonilla (*Sacando consecuencias: Una filosofía para el siglo XXI, Contra apocalípticos: Ecologismo, Animalismo, Posthumanism*o), se ocupa de la cuestión del yo (y más) con el objetivo de ahorrarnos la ver-

güenza de incurrir en malentendidos como el que acabamos de mencionar. Dividido en seis capítulos (sin contar la introducción y el epílogo) se propone situarnos justo en el centro del debate moderno sobre la subjetividad, el alma, la libertad, el libre albedrío y la consciencia, que en nuestros días está más vivo que nunca y no deja de sorprendernos con hipótesis (y demostraciones) que desafían los límites del no siempre sano sentido común.

Si no la he entendido mal, la tesis principal con que Zamora inaugura su ensayo expresa que la realidad experimentada no es, valga la redundancia, menos real por el hecho de que sea producto de intrincadas maniobras cerebrales (contra quienes ponen en duda esto, por cierto, incluso un idealista como Schopenhauer dirá que están para que los encierren). La sencillez y evidencia de esta proposición, sin embargo, contrasta con la poca aceptación que ha tenido por parte de pensadores de todos los tiempos, quienes la han retorcido hasta el dislate. Alrededor del concepto de «subjetividad» orbitan las más grandes anfibologías que pueda imaginarse.

En la introducción, Zamora advierte que la constitución ficcional del yo, en el sentido que Kant otorga a su noción de «apercepción trascendental», no procede de la actividad de un absoluto metafísico, como han querido ver los idealistas (sobre todo alemanes), sino de un montón de estímulos que excitan nuestras neuronas, cuya organización en patrones se convierte en todas aquellas sensaciones que percibimos y que sentimos *nuestras*. Aun así, no se olvida de que no son lo mismo las imágenes que nos llegan de los objetos (fenómenos), que los objetos considerados en sí mismos (noúmenos, entendidos como *Grenzbegriff,* conceptos límite). Pero aunque los objetos existan *fuera* de nosotros, realidad (o lo que nosotros entendemos por tal concepto) sólo hay una: la representada (por suerte). Y es que cumple en el entramado orgánico una función muy clara:

Al fin y al cabo, la función biológica primordial de los cerebros, o de los sistemas nerviosos en general, no solo en el caso de los seres humanos, es contribuir a que el individuo desarrolle una conducta exitosa (lo que, en términos biológicos, equivale a vivir lo bastante como para poder dejar muchos descendientes).

En el capítulo «El mito del espíritu», se despacha brevemente la cuestión del alma, recordándonos que su creencia inicialmente no trajo aparejada la convicción en su inmortalidad o persistencia en el más allá, y con ello da paso a la controversia generada en torno a la conexión existente entre mente y organismo, tan controvertida a lo largo de la historia de la filosofía. Pasa revista a muchas de las «experiencias extrañas» que a menudo aducen quienes defienden que existe una escisión entre ambas facultades, tales como las experiencias extracorporales, las cercanas a la muerte o las místicas. Las primeras han recibido la atención de numerosos estudios que las interpretan como percepciones distorsionadas, producto de aquellas áreas cerebrales encargadas de reconstruir el entorno y nuestra auto-percepción al margen de toda intencionalidad (de forma inconsciente), y cuya persuasiva sensación de realidad debe entenderse como *marca de la casa* (lo propio de una alucinación es que parezca real).

Es el caso de la ilusión de la mano de caucho. *En este experimento, un sujeto extiende ambas manos sobre una mesa. La mano derecha se oculta de su vista mediante una pantalla, al lado de la cual se coloca una mano de caucho que sí que puede ver. El experimentador va tocando simultáneamente la mano real del sujeto (que está*

oculta de su vista) y la mano de caucho, haciendo exactamente los mismos contactos en ambas. Al cabo de poco tiempo, el sujeto empieza a percibir que la mano de caucho es su mano real, e incluso puede llegar a percibir un «brazo fantasma» que conecta esa mano con su hombro.

De las segundas destaca que los signos distintivos que se les suelen atribuir, como la contemplación de un túnel oscuro o la de un ser luminoso, no son tan reseñables como podría pensarse, porque en realidad se dan con poca frecuencia y, además, han sido replicados con notable éxito en el caso de «los estudios sobre *pérdida de consciencia inducida por aceleración,* un fenómeno habitual entre los aviadores de combate, astronautas y pilotos de fórmula 1». Por lo general, se deben a una falta de riego sanguíneo y a la carencia de oxígeno en el sistema límbico, responsable de las emociones y culpable de la «sensación de dicha y profunda tranquilidad que suele acompañar a estas experiencias». Por último, las terceras, conocidas por ser inefables, profundas, fugaces y pasivas, no serían propiamente alucinaciones, sino estados de consciencia más primitivos, cercanos a los de los animales.

En «El mito del ordenador», Zamora se ocupa de la interesantísima cuestión, muy en boga en nuestros días, de si se podría establecer una correlación entre nuestro cerebro y un ordenador, y si, en fin, una máquina puede llegar a pensar. El óbice, señala nuestro autor, con que se tropiezan quienes se posicionan a favor de esta posibilidad (los llamados computacionistas) reside en la imposibilidad de probar que, efectivamente, una máquina está pensando (en sí) y no realizando un simulacro de pensamiento, que sólo parezca pensar *de cara a la galería* (para sí), sin que en su *interior* haya nada remotamente parecido a un pensamiento. Esto, es cierto, no cerraría la puerta a que una parte del desempeño intuitivo del pensamiento, en el futuro, pudiera llegar a reducirse a una serie de «algoritmos o programas rutinarios», pero en tal caso, dichos «programas deben de ser de un tipo radicalmente distinto al de los programas típicos de la IA contemporánea».

Zamora se adhiere a la crítica que John Searle presenta en *Mentes, cerebros y ciencia* (1984):

Imaginemos que el test de Turing se lleva a cabo en chino (si sabes chino, cámbialo por cualquier idioma que no entiendas), y que, en vez de haber un ordenador dentro de la habitación, estás tú, con un enorme libro o enciclopedia donde aparecen escritas todas las instrucciones que componen el software con el que íbamos a programar el ordenador. Cuando te llega una pregunta, buscas en las instrucciones los pasos que el programa dice que hay que dar para hallar la respuesta, sigues esos pasos, escribes la solución y la envías. Supongamos que el programa es exitoso. [...] ¿Significa eso que la máquina entiende chino? Según Searle, parece obvio que no, porque tú no has entendido en absoluto las preguntas ni las respuestas, es decir, puedes haber hecho todo el proceso sin entender nada, solo aplicando mecánicamente las reglas que hay en el libro.

El conexionismo tampoco está exento de problemas. Este enfoque, centrado no tanto en los algoritmos, como en la «arquitectura de las propias redes neuronales», no define muy bien qué cabría esperar tras el perseguido mapeo, completo y detallado, de las neuronas y las sinapsis, ya que efectivamente se dispondría del mapa, pero no de las claves exigidas para su desciframiento.

Así, la esperanza del transhumanismo, consistente en poder subir la mente a la nube o almacenarla en un disco duro, parece

quedar truncada. Además, Zamora opone los siguientes motivos. Por un lado, resulta imposible (al menos con los medios actuales y seguramente con los que tendremos a medio plazo) extraer la descomunal cantidad de datos de un cerebro sin dañarlos o alterarlos introduciendo en ellos elementos intrusivos. Y, por otro, «ignoramos por completo el código (o mejor dicho, la suma de miles o millones de códigos) que convierte esos datos en recuerdos», ya que la mente no se produce con el simple amontonamiento de datos, sino que algún tipo de proceso *trasero* debe integrarlos y transformarlos en estados mentales (todavía nadie ha visto que al llenarse el disco duro de su ordenador haya cobrado conciencia de sí).

Aparte de Searle, muchos otros nombres de festejados filósofos y neurofisiólogos desfilan por sus páginas. Leemos, por ejemplo, que David Chalmers, en *La mente consciente* (1996) presenta la conocida hipótesis (o más bien el experimento mental) de que puedan «existir cuerpos sin mente, seres humanos que carezcan de una mente consciente, a pesar de que todo su comportamiento sea exactamente como el de una persona *normal*». Dado que el comportamiento es el resultado de fuerzas físicas que operan sobre células, moléculas y átomos que componen a los individuos, podría darse el caso de que la interacción entre estos elementos fuese suficiente para explicarlo, sin el adjunto concurso de lo mental. Así pues, Chalmers extrae la conclusión de que las propiedades físicas y mentales de «un organismo son propiedades diferentes, que pueden darse por separado». Así que, en principio, «podrían existir unas al margen de las otras». La objeción de Zamora reside en preguntarse por qué, si los que carecen de mente y los que sí la tienen poseen procesos neurológicos similares o idénticos, en un caso no generan *qualia* (que es, por ejemplo, la experiencia mental subjetiva del amarillo al mirar un limón) y en el otro sí.

En el capítulo «Ciencia y consciencia», nos presenta dos teorías que representan los «principales enfoques científicos sobre el problema mente-cuerpo». Una pertenece a Giulio Tononi, médico psiquiatra y neurocientífico italiano, que resume las cualidades esenciales de la consciencia en cinco puntos (existencia intrínseca, composición, información, integración y exclusión) y le sirven para establecer cuáles deben ser las exigencias mínimas que debe cumplir un sistema físico que aspire a convertirse en soporte de experiencia.

Según Tononi, poseer experiencia consistiría en ser un sistema con un valor de Φ mayor que 0, mientras que cuál es la experiencia que se está teniendo, en qué consiste experimentar esa experiencia (o sea, lo que en el capítulo anterior llamábamos qualia), viene dado por la información específica contenida por el sistema en ese momento.

Para Zamora, no obstante, Tononi se equivoca con su teoría de la consciencia como «información integrada» al creer que lo cualitativo de los *qualia* tiene que ver con la diferencia entre cantidad y cualidad, puesto que en realidad «la diferencia más relevante en este caso es la que hay entre lo *formal* (la estructura) y lo *material* (qué materia o sustancia es la que posee dicha estructura)», y pone un ejemplo:

Puesto que los sonidos y los colores dependen de sendos tipos de vibraciones (mecánicas o electromagnéticas), es posible que entre dos sonidos y entre dos colores haya una misma relación estructural (digamos, por ejemplo, que puede ocurrir en ambos casos que la frecuencia con la que vibra el estímulo que nos hace percibir cierto sonido o un color sea el 90% de la frecuencia con la que vibra otro estímulo). Pero esta relación estructural está total-

mente oculta para nosotros al percibir sensorialmente esos colores y sonidos: aunque los sonidos nos parecen distintos el uno del otro, y lo mismo los colores, no percibimos en absoluto que la relación entre los dos sonidos (uno ligeramente más agudo que el otro) tenga nada que ver, ninguna semejanza, con la relación entre los dos colores (uno más rojo, otro más azul, por ejemplo).

La otra teoría que aparece en este capítulo, de ribetes biológicos, corresponde al «espacio global de trabajo», formulada por Bernard Baars en los años 80 del pasado siglo. Se atreve a dar respuesta a cuestiones que la de la «información integrada» omite quizá deliberadamente por impotencia, a saber: por qué los procesos cerebrales son tan complejos y por qué muchos de ellos (de hecho, la mayoría) son inconscientes, dejando para lo consciente tan estrecho margen de maniobra.

La idea del espacio de trabajo global» es una respuesta a estas cuestiones: el principal papel de la consciencia consistiría en lograr que algunos ítems de toda esa gran cantidad de información que el cerebro está procesando de manera mecánica» o inconsciente puedan estar disponibles para ser utilizados por cualquier otro subsistema cerebral. [...] Dicho de otro modo, la consciencia habría evolucionado como una estrategia para mitigar los efectos de una excesiva modularidad cerebral.

Uno de los mayores valedores de esta teoría, señala Zamora, es Stanislas Dehanene, para quien la dificultad de definir los *qualia*, cuya naturaleza sólo comprendemos de forma subjetiva e intuitiva, resulta del simple «hecho de que la cantidad de información sensorial de una experiencia consciente es demasiado grande como para que podamos ponerla en palabras con los recursos que caben a través del *cuello de botella* de nuestra capacidad de atención consciente».

Por último, Zamora actualiza con nuevos detalles el inveterado debate entre determinismo y libre albedrío. Niega al yo la capacidad de operar sobre las cosas de manera incondicionada, pues en realidad la sensación de subjetividad, cuyo signo distintivo consiste en que parezca que todo desemboca en una experiencia unitaria llamada «yo», es también producto del movimiento de neuronas activándose *hic et nunc*. Asimismo, rechaza que pueda darse una posibilidad alternativa realmente existente, esto es, que algo que hemos hecho de una determinada forma pueda haber sido de otro modo, porque a la postre no existe manera de demostrarlo más que en la ficción, *a posteriori* (o como dice Schopenhauer en *Sobre la voluntad en la naturaleza*: se puede querer algo, pero no querer querer algo). La sensación de que en realidad podríamos haber elegido lo que al final no elegimos es, pues, una ilusión de libertad.

Acaba *En busca del yo. El mito del sujeto y libre albedrío* con un comentario al emergentismo, en concreto la propuesta del filósofo alemán Christian List. Según este autor, la realidad se encuentra compartimentada en diversos niveles emergentes regidos por leyes autónomas. Los principios de un nivel no funcionan en los otros. Por ejemplo, todo lo que sabemos sobre partículas subatómicas no nos sirve para entender cómo funcionan las flores. Cada nivel, a su vez, supone un peldaño arriba en la escala ontológica. Una bacteria está por encima de un *quark*. Un estómago está por encima de una bacteria. Y un ser humano está por encima de un estómago. Por eso, al decir de List, el estado neurológico de la intencionalidad humana no puede rastrearse causalmente (que hoy decida quedarme en casa leyendo un libro en vez de irme a correr

no puede explicarse mediante la interacción mecánica de los átomos que conforman mi cerebro, puesto que la motivación reside en una dimensión aparte compuesta por unidades mínimas distintas). Las decisiones, así, no son «meras ficciones epifenoménicas», sino realmente libres, pues no están sujetas a las leyes deterministas que imperan en los niveles ontológicos inferiores. La réplica de Zamora nos parece muy apropiada, ya que List se ve en serios apuros para conciliar dos aspectos tan contradictorios de su teoría:

Si la naturaleza es determinista en su nivel más fundamental, entonces solo existe una cadena posible de estados a nivel micro (a saber, la cadena de micro-estados que efectivamente ocurre en ese nivel), y, de manera correspondiente, solo existe una cadena posible de macro-estados a cualquier macro-nivel superior: los macro-estados que vienen determinados por los micro-estados correspondientes.

Por todo lo dicho, *En busca del yo. El mito del sujeto y el libre albedrío* se muestra como una provechosa lectura que, en un estilo ameno y en ocasiones hasta humorístico, nos situará en el rastro de muchas de las teorías sobre filosofía de la mente y ontología que las cabezas más ingeniosas y extrañas han ido poniendo sobre la mesa en estos últimos años.

José Carlos Ibarra Cuchillo

NORMAS DE PUBLICACIÓN

La finalidad de *Daimon - Revista Internacional de Filosofía* es publicar trabajos de investigación en filosofía. *Daimon* es, desde 2001, una publicación cuatrimestral. Algunos de los números son monográficos y otros no. Los números monográficos son anunciados con antelación suficiente (al menos un año) mediante la correspondiente *llamada para aportaciones (call for papers),* en la que se anuncia el tema del monográfico y el nombre de la persona encargada de coordinarlo. En el caso de que un monográfico no reciba originales suficientes para completar el volumen (actualmente tenemos fijado un límite en torno a las doscientas páginas), se completará con una sección de artículos variados.

Formato de los originales: Véase en https://revistas.um.es/daimon/about/submissions

El texto de los artículos y de notas críticas que sea enviado para revisión NO debe contener datos personales del autor o autores, ni en el propio texto, ni en las propiedades del archivo informático, ni en las citas bibliográficas (en este último caso, cada cita de trabajos del autor ha de ser sustituida por la palabra "Autor" y el año de la publicación referida).
Las citas bibliográficas han de hacerse de acuerdo con el ESTILO APA a partir de *Publication Manual of the American Psychological Association*, *7th edition*, de 2020 (https://apastyle.apa.org/style-grammar-guidelines/index). Resumen en español de la 7ª ed. de estas normas en http://www.um.es/analesps/informes/APA7ed-resumenNormas-v10febr2021.pdf.

Procedimiento: Véase en http://revistas.um.es/index.php/daimon/about/submissions

Daimon. Revista Internacional de Filosofía

Publicación cuatrimestral. Número 93. Septiembre-Diciembre 2024

editum
EDICIÓN DE LA UNIVERSIDAD DE MURCIA

2023 CALIDAD REVISTAS CIENTÍFICAS ESPAÑOLAS FECYT