



# **UNIVERSIDAD DE BUENOS AIRES**

Facultad de Ciencias Exactas y Naturales,  
Departamento de Física

## **El rol de las interacciones sociales en la sabiduría de las multitudes y la dinámica de opiniones**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos  
Aires en el área de Ciencias Físicas

**Lic. Federico Ulises Barrera Lemarchand**

Director de Tesis: Dr. Joaquín Navajas

Co-Director de Tesis: Dr. Pablo Balenzuela

Consejero de Estudios: Dr. Diego Shalom

Lugar de trabajo: Laboratorio de Neurociencia, Universidad Torcuato di Tella

Fecha de defensa: 28 de agosto de 2024

Buenos Aires, 2024

# **El Rol de las Interacciones Sociales en la Sabiduría de las Multitudes y la Dinámica de Opiniones**

## **Resumen**

La toma de decisiones grupales constituye un área de investigación de gran relevancia para diversas áreas del conocimiento, como la Psicología Experimental y la Física Estadística. Existen diferentes abordajes y objetivos que intervienen en su estudio, como, por ejemplo, identificar métodos óptimos para la obtención de conocimiento y resolución de problemas de una multitud; resolver problemas en grupo por medio de estrategias eficientes de interacción; o generar consenso en problemas para los que no necesariamente existe una respuesta correcta. En todos los casos, se trata de contextos muy relevantes para las decisiones democráticas. En esta tesis se presentan tres abordajes diseñados para mejorar las chances de alcanzar consensos grupales, y extraer información útil para la resolución de problemas. En el Capítulo 2 se exhibe un método para mejorar la sabiduría de las multitudes en casos en los que suele fallar, mediante un incremento de la diversidad de estimativos de la multitud. En el Capítulo 3 se estudia qué estrategias utilizan grupos de discusión para alcanzar una respuesta consensuada en problemas de estimación, dando como resultado valores que superan ampliamente la sabiduría de las multitudes clásica. En el Capítulo 4 se desarrolla un modelo de formación de opinión basado en agentes, que incorpora un intercambio explícito de argumentos, y considera el impacto de sesgos cognitivos comunes, como el sesgo de confirmación y la homofilia. Hoy en día, cuando las opiniones políticas tienden a polarizarse, dilucidar este tipo de interrogantes se torna especialmente importante, tanto para fines teóricos (como el estudio del valor epistémico de la democracia) como prácticos (optimizar las discusiones y alcanzar consensos).

**Palabras clave:** sabiduría de las multitudes, sesgo de anclaje, teorema de diversidad predictiva, extremización, método de Fermi, interacciones grupales, sociofísica, modelos de agentes, dinámica de opiniones, teoría de argumentos persuasivos, sesgo de confirmación, homofilia, ciencias sociales computacionales

# The Role of Social Interactions on the Wisdom of Crowds and Opinion Dynamics

## Abstract

Group decision-making constitutes a highly relevant area of research across various fields of knowledge, such as Experimental Psychology and Statistical Physics. There are different approaches and objectives involved in its study, such as identifying optimal methods for acquiring knowledge and solving problems collectively; solving problems in groups through efficient interaction strategies; or generating consensus on problems where there may not necessarily be a correct answer. In all cases, these are highly relevant contexts for democratic decision-making. This thesis presents three approaches designed to enhance the likelihood of reaching group consensus and extracting useful information for problem solving. Chapter 2 showcases a method to improve the wisdom of crowds in cases where it tends to falter, by increasing the diversity of crowd estimates. Chapter 3 examines the strategies employed by discussion groups to reach a consensus on estimation problems, resulting in values that considerably surpass classical crowd wisdom. Chapter 4 develops an opinion formation agent-based model, incorporating explicit argument exchange and considering the impact of common cognitive biases such as confirmation bias and homophily. Nowadays, as political opinions tend to polarize, elucidating these kinds of questions becomes especially important, both for theoretical purposes (such as studying the epistemic value of democracy) and practical ones (optimizing discussions and reaching consensus).

**Keywords:** wisdom of crowds, anchoring bias, diversity prediction theorem, extremization, Fermi method, group interactions, sociophysics, agent-based models, opinion dynamics, persuasive arguments theory, confirmation bias, homophily, computational social science

## **Agradecimientos**

A Joaquín Navajas.

A Pablo Balenzuela.

A Viki, Nuri, Juli, Cande y Candy.

A Mariano Sigman, Julia Hermida, Cecilia Calero, Ariel Haimovici, María Luz Gadea, Andrea Goldin, Carolina Gattei, Diego de la Hera, Luciana Lucchina, Juliana Leone, Lucía Freira, Andrés Rieznik, Facundo Álvarez Hedúán, y Agustín Gravano. A Federico Zimmerman, Martina Tokatlian, Melina Vladisauskas, Guadalupe Rodríguez Ferrante, Analí Taboh, Milagros Azcueta, Martina Boscolo, Julieta Goldstein, Florencia Lee, Santos Espina Mairal, Alexia Aquino, Andrés Cotton, Belén Zanoni, Leonardo Svarc, Antonella Furchi, Patricio Peralta, y Luz Alba Posse.

A Diego Shalom y Mariano Mayochi.

A Gabi y Jaz.

A Cami, Robin, Fran y Taiel.

A Jessie, Nico, Giu, Ivo, Christian.

A Facu, Juan, Maia, y Maru.

A Julián, Mery, Agus.

A María Luz. A María.

A Rosario, A Gastón.

A Verónica y Marcelo.

A Susana.

A Guillermo y Viviana.

# Publicaciones Asociadas con esta Tesis

## En revistas

- Barrera-Lemarchand, F., Lescano-Charreau, V., Ruiz, J., Cáceres, N., Carrillo, F., Navajas, J. Collective Wisdom and the Fermi Method: Improving the Accuracy of Deliberative Groups. En escritura.
- Barrera-Lemarchand, F., Balenzuela, P., Bahrami, B., Deroy, O., & Navajas, J. (2024). Promoting Erroneous Divergent Opinions Increases the Wisdom of Crowds. *Psychological Science*,  
<https://doi.org/10.1177/09567976241252138>
- Röseler, L., Weber, L., Stich, E., Günther, M., Helgerth, K. A. C., Wagner, F. S., Antunovic, M., Bahník, S., Barrera-Lemarchand, F., ... & Schütz, A. (2022). The open anchoring quest dataset: Anchored estimates from 96 studies on anchoring effects. *Journal of Open Psychology Data*, 10(1), 16.  
<http://doi.org/10.5334/jopd.67>
- Barrera Lemarchand, F., Semeshenko, V., Navajas, J., & Balenzuela, P. (2020). Polarizing crowds: Consensus and bipolarization in a persuasive arguments model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(6).  
<https://doi.org/10.1063/5.0004504>

## En proceedings:

- Barrera-Lemarchand, F., Lescano-Charreau, V., Ruiz, J., Cáceres, N., Carrillo, F., Navajas, J. Collective Wisdom and the Fermi Method: Improving the Accuracy of Deliberative Groups. *SCT Proceedings in Interdisciplinary Insights and Innovations*.  
<https://doi.org/10.56294/piii2024257>

## Presentaciones en Congresos:

- *Sabiduría Colectiva y el Método de Fermi: Mejorando la Exactitud de Grupos Deliberativos.* **Barrera-Lemarchand, F.**; Lescano-Charreau, V.; Ruiz, J.; Cáceres, N; Carrillo, F; Navajas, J. **Charla Corta.** V Jornadas de Fundamentos y Aplicaciones de la Interdisciplina, en la Universidad Nacional de Luján (sede CABA), CABA, Argentina (2024).
- *Extremizing Judgements Produces More Inaccurate Individuals but Wiser Crowds.* **Barrera-Lemarchand, F.**; Navajas, J. **Presentación de Póster.** The Cognitive Science Society Conference (CogSci) 2023, en el International Convention Centre Sydney, Sydney, Australia (2023).

- *Linguistic Complexity Reduces the Perceived Validity of Moral Arguments.* **Barrera-Lemarchand**, F.; Navajas, J. **Parallel Talk.** 9th International Conference on Computational Social Science (IC<sub>2</sub>S<sub>2</sub>), en la University of Copenhagen, Copenhagen, Denmark (2023).
- *¿Menos es Más? El Impacto de la Complejidad Discursiva en la Valoración de Argumentos Morales,* **Barrera-Lemarchand**, F.; Navajas, J. **Charla Corta.** IV Jornadas de Fundamentos y Aplicaciones de la Interdisciplina, en la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. CABA, Argentina (2023). <https://doi.org/10.56294/sctconf202347>
- *Polarización e Inteligencia Colectiva.* **Barrera Lemarchand**, F.; Navajas, J. **Charla Simposio.** XXVII Jornadas de Epistemología de las Ciencias Económicas, Facultad de Ciencias Económicas, Universidad de Buenos Aires, CABA, Argentina (2021).
- *Multitudes Polarizadas: Consenso y Bipolarización en un Modelo de Argumentos Persuasivos.* **Barrera Lemarchand**, F.; Semeshenko, V.; Navajas, J.; Balenzuela, P. **Charla Corta.** XVIII Taller Regional de Física Estadística y Aplicaciones a la Materia Condensada en la FAMAF, Universidad Nacional de Córdoba, Córdoba, Argentina (2021).
- *Cómo aprovechar el sesgo de anclaje para reducir el error colectivo de una multitud.* **Barrera Lemarchand**, F.; Navajas, J. **Presentación de Póster.** II Jornadas de Fundamentos y Aplicaciones de la Interdisciplina, en la Facultad de Ciencias Exactas y Naturales (virtual), Universidad de Buenos Aires. CABA, Argentina (2021).
- *Sabiduría de las multitudes y sesgo de anclaje: cómo reducir el error colectivo de una multitud.* **Barrera Lemarchand**, F.; Navajas, J. **Presentación de Póster.** XVII Reunión Nacional y VI Reunión Internacional de la Asociación Argentina de Ciencias del Comportamiento (AACC). Posadas, Misiones, Argentina (2019).
- *El laboratorio de la percepción: una experiencia de extensión e investigación extra-académica, horizontal e interdisciplinaria.* J. Sánchez Beisel, M.; Merlo, S. A.; **Barrera Lemarchand**, F.; Ramos, M. S.; Christofalikis, C.; Masia, F. ; Corti, J. F.; Cossio Mercado, C.; Caro, I.; Fernández Larrosa, P. N. (todos los autores contribuyeron equitativamente al trabajo). **Presentación de Póster.** VII Jornadas de Investigación de la Facultad de Psicología y Psicopedagogía, Universidad del Salvador. CABA, Argentina (2019).
- *El laboratorio de la percepción: una experiencia de extensión e investigación extra-académica, horizontal e interdisciplinaria.* J. Sánchez Beisel, M.; Merlo, S. A.; **Barrera Lemarchand**, F.; Ramos, M. S.; Christofalikis, C.; Masia, F. ; Corti, J. F.; Cossio Mercado, C.; Caro, I.; Fernández Larrosa, P. N. (todos los autores contribuyeron equitativamente al trabajo). **Presentación de Póster.** Jornadas de Fundamentos y Aplicaciones de la Interdisciplina, en la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. CABA, Argentina (2019).

# Índice general

<b>1. Introducción .....</b>	10
<b>1.1. Problemas Fácticos .....</b>	10
<b>1.2. Problemas de Opinión .....</b>	14
<b>1.3. Metodologías.....</b>	16
<b>1.4. Estructura.....</b>	19
<b>2. Promover opiniones erróneas divergentes aumenta la sabiduría de las multitudes .....</b>	22
<b>2.1. Resumen.....</b>	22
<b>2.2. Introducción .....</b>	23
<b>2.3. Aumento de la Sabiduría de las Multitudes mediante Extremización .....</b>	25
<b>2.4. Modelo.....</b>	28
<b>2.5. Experimento 1: estimación de cantidades no acotadas.....</b>	31
<b>2.5.1 Métodos.....</b>	31
<b>2.5.2 Resultados y Discusión.....</b>	33
<b>2.6 Experimento 2: estimación de cantidades acotadas con anclas fijas .....</b>	34
<b>2.6.1 Métodos.....</b>	34
<b>2.6.2 Resultados y Discusión.....</b>	37
<b>2.7 Experimento 3: predicción .....</b>	41
<b>2.7.1 Métodos.....</b>	41
<b>2.7.2 Resultados y Discusión.....</b>	44
<b>2.8 Experimento 4: anclas auto-generadas .....</b>	46
<b>2.8.1 Métodos.....</b>	46
<b>2.8.2 Resultados y Discusión.....</b>	48
<b>2.9 Discusión General.....</b>	50
<b>3. La Implementación Colectiva del Método de Fermi Subyace a la Sabiduría de Multitudes Deliberativas.....</b>	54
<b>3.1. Resumen.....</b>	54
<b>3.2. Introducción .....</b>	55
<b>3.3. Resultados.....</b>	58
<b>3.3.1. Experimento 1 .....</b>	58

<b>3.3.2. Experimento 2 .....</b>	67
<b>3.3.3. Experimento 3 .....</b>	72
<b>3.4. Discusión General .....</b>	74
<b>3.5. Métodos.....</b>	78
<b>3.5.1. Desarrollo de una aplicación web para experimentos de decisiones grupales.....</b>	78
<b>3.5.2. Participantes y Preguntas.....</b>	79
<b>3.5.3. Procedimiento.....</b>	82
<b>3.5.4. Exclusiones.....</b>	83
<b>4. Multitudes Polarizadas: Consenso y Bipolarización en un Modelo de Argumentos Persuasivos .</b>	84
<b>4.1. Resumen.....</b>	84
<b>4.2. Introducción .....</b>	85
<b>4.3. Modelo.....</b>	89
<b>4.3.1. Estado de los Agentes.....</b>	89
<b>4.3.2. Dinámica .....</b>	91
<b>4.4. Resultados.....</b>	94
<b>4.4.1. Modelo Imparcial.....</b>	94
<b>4.4.2. Modelo con Sesgo de Confirmación .....</b>	98
<b>4.4.3. Modelo con Homofilia.....</b>	102
<b>4.5. Discusión General .....</b>	104
<b>5. Discusión General .....</b>	108
<b>6. Bibliografía .....</b>	111
<b>7. Apéndice 1.....</b>	125
<b>Teorema de Diversidad Predictiva .....</b>	125
<b>Condiciones para Sobrepasar la Sabiduría de las Multitudes .....</b>	126
<b>Competencia entre Aumentar Diversidad y Exactitud.....</b>	129
<b>Evaluación de Diferentes Métodos de Agregación .....</b>	130
<b>Anclas Auto-Generadas y la Sabiduría de la Multitud Interna.....</b>	134
<b>Definición del “Índice de Anclaje” .....</b>	135
<b>Declaración de Prácticas Abiertas.....</b>	135
<b>Variables de Diseños Experimentales .....</b>	136
<b>Experimento 1: preguntas y respuestas correctas .....</b>	137
<b>Experimento 2: preguntas y respuestas correctas .....</b>	138
<b>Experimento 3: preguntas y respuestas correctas.....</b>	140

<b>Experimento 4: preguntas y respuestas correctas.....</b>	140
<b>8. Apéndice 2.....</b>	141
<b>Preguntas realizadas en los distintos experimentos del Capítulo 3.....</b>	141
<b>Figura Suplementaria 1.....</b>	142
<b>Figura Suplementaria 2.....</b>	143
<b>9. Apéndice 3.....</b>	144

# Capítulo 1

## 1. Introducción

La mayoría de las personas de nuestra sociedad se ve envuelta diariamente en discusiones que derivan en distintos tipos de decisiones. Debates políticos (como los debates presidenciales), referendos (como el del "Brexit" de 2016), grupos de discusión (paneles médicos, jurados, etc.) e incluso reuniones laborales, son ejemplos de discusiones – en escalas muy diferentes –, en las que se deben tomar decisiones importantes. Algunas de estas decisiones pueden estar relacionadas con la resolución de problemas *fácticos o intelectivos*: aquellos que tienen una respuesta correcta, y que en general involucran estimar magnitudes desconocidas. En otros casos, se trata de problemas *de opinión o no intelectivos*: sin una respuesta correcta clara, pero que pueden ser igual o más relevantes para nuestra sociedad. Quizás el ejemplo más sobresaliente de este tipo de decisiones lo constituyen las elecciones democráticas.

### 1.1. Problemas Fácticos

Sir Francis Galton, un aclamado científico del siglo XIX conocido por sus aportes en matemática, física de la atmósfera, psicología, y biología, entre otras áreas, buscaba demostrar que las decisiones democráticas no pueden derivar en resultados útiles para la sociedad. Razonaba que, si uno agrega información errónea proveniente de muchas fuentes, el resultado final no podría nunca ser información útil o relevante (Surowiecki, 2005). En 1906, decidió hacer un estudio para corroborar su intuición: en una feria local se ofrecía un buey como premio a aquella persona que

pudiera proveer el estimativo más exacto del peso del mismo. Galton anotó los estimativos de los más de cuatrocientos participantes, de los cuales la gran mayoría carecía por completo de conocimientos previos que pudieran aprovechar para resolver esa tarea. Tras agregar los estimativos individuales de todos ellos, el resultado presentaba una exactitud llamativa –mayor a la de cualquiera de aquellos estimativos (Galton, 1907). Esto llevó a Galton a concluir que la democracia podía tener cierto sustento empírico gracias a este fenómeno, el cual hoy en día se conoce bajo el nombre de “Sabiduría de las Multitudes” (De Condorcet, 1785; Surowiecki, 2005).

En la actualidad, el efecto de Sabiduría de las Multitudes es empleado para resolver problemas de estimación en numerosas aplicaciones de gran relevancia, como mejorar los diagnósticos médicos (Kurvers et al., 2016), prever eventos geopolíticos (Mellers et al., 2014), predecir los mercados financieros (Ray, 2010), realizar ingeniería inversa del olor de moléculas (Keller et al., 2017), y verificar hechos en las noticias (Allen et al., 2021).

La contracara de este fenómeno es que “no siempre funciona”. En algunas circunstancias, se produce lo que es conocido como “Locura de las Multitudes” (MacKay, 1841): la convergencia de un gran número de individuos a creencias y decisiones irracionales. Cuando esto ocurre, una multitud puede coincidir en la propagación de creencias incorrectas (Friedkin et al, 2016) o no sustentadas en evidencia (Festinger, 1962). Ejemplos de este fenómeno incluyen las burbujas financieras (Chari & Kehoe, 2004) y el fundamentalismo (Festinger et al, 1956).

Dada su relevancia práctica, comprender las condiciones bajo las cuales las multitudes producen estimaciones exactas se ha convertido en un tema cada vez más preponderante en las ciencias psicológicas (Kameda, Toyokawa, y Tindale, 2022; Karachiwalla y Pinkow, 2021; Navajas et al., 2018; Kao y Couzin, 2014).

Un impulsor importante de la exactitud colectiva es la diversidad de opiniones en la multitud (Hong y Page, 2004; Page, 2008; Becker, Porter y Centola, 2019; Shi et al., 2019; Jönsson, Hahn y Olsson, 2015). Una intuición simple subyace a esta afirmación: cuando las multitudes producen estimaciones diversas, es probable que algunos individuos subestimen la respuesta correcta, mientras que otros la sobreestimen. Por lo tanto, cuanto más diversa sea la multitud, mayor será la probabilidad de que los errores individuales se cancelen en el proceso de agregación.

En el Capítulo 2 de esta tesis se presenta un método novedoso diseñado para aumentar la diversidad de estimativos en una multitud, permitiendo mejorar la sabiduría de las multitudes en aquellos casos en los que “no funciona” (y donde, por ende, existe la posibilidad de mejorar). Un modelo matemático similar a un promedio pesado con pesos no lineales, y cuatro estudios experimentales (tres de ellos pre-registrados) proveerán un marco teórico sólido y evidencia empírica abundante, a favor del empleo de este método.

Asimismo, muchas investigaciones previas han explorado el impacto de la influencia social en la sabiduría de las multitudes, arrojando resultados disímiles. Mientras que algunos estudios han informado efectos adversos de la influencia social (Raafat, Chater y Frith, 2007; Lorenz et al., 2011; Madirolas y de Polavieja, 2015), otros han demostrado que puede mejorar la exactitud colectiva (Gürçay, Mellers y Baron., 2015; Mellers, 2014; Bahrami, 2010; Juni y Eckstein, 2015). Por ejemplo, un estudio previo de este grupo de investigación ha evidenciado que la agregación de estimaciones colectivas de unos pocos grupos interactuantes puede superar ampliamente la sabiduría de muchos individuos (Navajas et al., 2018). Aunque este efecto ha sido ampliamente replicado en diversos entornos, los procedimientos empleados por dichos grupos para llegar a estimaciones colectivas tan exactas aún restan por dilucidarse.

Una posibilidad, respaldada por la literatura previa, consiste en recalcular la variable de interés dividiendo el problema principal en problemas de estimación más pequeños, y realizando los cálculos necesarios para llegar a la respuesta final. Por ejemplo, para estimar el número de escalones en la escalera de un edificio grande, se podría suponer un número dado de pisos, una altura específica de un escalón individual, y una cierta cantidad de escalones en cada piso. Al combinar estas estimaciones independientes, es posible llegar a una estimación final del número total de escalones. Esta estrategia, comúnmente conocida como el "Método de Fermi" (en honor a Enrico Fermi, quien la promovía y empleaba para la estimación del orden de magnitud de variables desconocidas), ha demostrado ser efectiva en una amplia variedad de ámbitos (Nityananda, 2014). Otra posibilidad es que los grupos utilicen algún método de agregación (como la media aritmética, la mediana, promedios pesados por confianza, etc.), para llegar a estos estimativos. Este método no involucra un razonamiento colectivo, sino la comunicación de los estimativos individuales previos, que luego son agregados.

En el capítulo 3 de este trabajo se analizará en detalle el mecanismo utilizado por grupos de discusión en experimentos similares al presentado en Navajas et al. (2018). Se estudiarán las conversaciones escritas de los participantes, que interactuaban en grupos de cuatro, en pos de resolver problemas simples de estimación numérica. Además de replicar los resultados de Navajas et al. (2018), se provee sustento a la hipótesis de que aquellos grupos con mayor exactitud realizan procedimientos de razonamiento colectivo similares al “Método de Fermi”, en lugar de utilizar métodos de agregación de sus estimativos individuales previos.

Ambos capítulos implican un avance sustancial hacia una mejora en la sabiduría de las multitudes, a la vez que contribuyen a responder interrogantes fundamentales sobre las variables que intervienen en este fenómeno.

## **1.2. Problemas de Opinión**

Si bien los problemas fácticos resultan ser extremadamente importantes en innumerables aplicaciones, y hasta se los ha usado como fundamento de los mecanismos democráticos de toma de decisiones, también es cierto que muchos de los problemas de la vida cotidiana no presentan una respuesta correcta clara, sino que se trata de problemas de opinión. Con frecuencia, en este tipo de interacciones, se invocan e intercambian argumentos, lo que eventualmente puede llevar a que se modifiquen las opiniones iniciales, así como las decisiones a tomar. Estas incidencias guiaron el desarrollo de numerosos estudios experimentales y teóricos, centrados en comprender la dinámica de formación de opinión. Desde un punto de vista teórico, la Física Estadística proporciona herramientas cuantitativas muy poderosas, particularmente útiles para el estudio de sistemas complejos compuestos por muchos agentes interactuantes –como se puede apreciar en varias revisiones recientes y exhaustivas (Baronchelli, 2018; Castellano et al., 2009; Helbing, 2010; Weidlich, 1991).

En este marco de referencia, los modelos basados en agentes constituyen una técnica ampliamente difundida para derivar estados macroscópicos a partir de interacciones microscópicas simples entre agentes interactuantes. Se han aplicado eficientemente para estudiar diversos fenómenos y entornos, como el comportamiento y el tráfico de peatones

(Castellano et al., 2009), la agricultura (Berger, 2001), la economía (Farmer & Foley, 2009), los sistemas socio-ecológicos (Filatova, 2013), la demografía (Billari & Prskawetz, 2012), entre muchos otros (Castellano et al., 2009; Goldstone & Janssen, 2005; Kohler & Gumerman, 2000).

La dinámica de opiniones produce una serie de estados macroscópicos que ocurren naturalmente en nuestra sociedad (Lorenz et al., 2020). Dada su relevancia política, económica y social, estos estados han sido estudiados ampliamente (Burnstein & Vinokur, 1977; Sunstein, 2002; Jager & Amblard, 2005; Sampedro & Pérez, 2008). El consenso (cuando todos los miembros del grupo adhieren a la misma opinión) y la bipolarización (cuando se forman dos grupos distintos y polarizados) son ejemplos relevantes que se encuentran comúnmente en temas con afirmaciones binarias (“a favor” o “en contra”, o una elección por ballottage, por ejemplo). Aunque la bipolarización puede ser inducida introduciendo un mecanismo de influencia negativa (distanciamiento intencional de agentes disímiles), la evidencia que respalda su existencia es insuficiente y controvertida (Krizan & Baron, 2007). Esto ha llevado a la propuesta de otros mecanismos capaces de reproducir dicho estado colectivo (Mas and & Flache, 2013). En Mas & Flache (2013), los autores proponen un modelo para alcanzar bipolarización en ausencia de influencia negativa, mediante la introducción de un intercambio explícito de argumentos, junto con la homofilia (mayor probabilidad de interacción con agentes similares en lugar de disímiles). Sus suposiciones se basan en la Teoría de Argumentos Persuasivos (PAT), que establece que los cambios de opinión se derivan de intercambios de argumentos, y dependen tanto del número, como de la fuerza (o persuasión) de los mismos.

En el capítulo 4 se presenta un modelo basado en agentes donde los estados de los agentes están representados por una opinión discreta, pero que gozan de posturas internas continuas

subyacentes (su opinión depende del valor de la postura, que se mueve en un rango acotado de posibles valores). Este modelo constituye una extensión de aquel presentado por Balenzuela et al. (2015); no obstante, a diferencia de este último trabajo, las interacciones se dan mediante un intercambio explícito de argumentos, basado en PAT. Se desarrolla un conjunto de reglas para este intercambio, y se explora el nuevo espacio de parámetros, introduciendo también modificaciones destinadas a modelar ciertos sesgos cognitivos reportados previamente en la literatura, que han sido vinculados a los estados macroscópicos observados en la sociedad – como el sesgo de confirmación, o la homofilia (Parmley, 2006; Lehner et al., 2008; Byrne, 1961; Moscovici & Zavalloni, 1969).

### **1.3. Metodologías**

Los métodos de investigación empleados en esta tesis, de carácter interdisciplinario, son propios de aquellas áreas que estudian cuantitativamente los sistemas sociales, como la sociofísica, la economía comportamental, la psicología cognitiva y social, y las ciencias sociales computacionales. Se trata de la utilización de un bagaje de herramientas estadísticas, de modelado matemático y computacional, y de experimentación con humanos, para dilucidar los mecanismos microscópicos de interacción que permiten entender fenómenos macroscópicos, obtener información útil de sistemas sociales para la resolución de problemas (fácticos y no fácticos), y realizar intervenciones para optimizar su funcionamiento.

La sociofísica es una disciplina que parte de la física estadística, y que centra su objeto de estudio en los sistemas sociales, y, en particular, en el comportamiento social humano y sus

dinámicas (Castellano et al., 2009; Galam, 2012; Haken, 1978; Helbing, 2010; Weidlich, 1991, 2000). Sus herramientas principales consisten en el modelado matemático y computacional, e incorpora elementos de áreas como redes complejas, los modelos de agentes, y la dinámica no lineal. Si bien la idea de modelar físicamente fenómenos sociales es casi tan antigua como la idea de modelar estadísticamente los fenómenos físicos (Ball, 2006; Majorana, 1942), los avances tecnológicos recientes –que permiten contar con grandes volúmenes de datos; experimentación online con cantidades de participantes muy superiores a los experimentos tradicionales; y acceso continuo y sencillo a dispositivos portátiles de medición, como los smartphones– generaron una explosión en la investigación cuantitativa y el interés por el modelado físico de dichos fenómenos (Conte et al., 2012; D. M. J. Lazer et al., 2020; D. Lazer et al., 2009). Tal es así que, en los últimos años, la idea de estudiar a la sociedad dentro del marco de la física estadística ha pasado de ser una declaración filosófica de principios a un esfuerzo de investigación concreto que involucra a un gran número de investigadores/as y una profusa producción científica (Castellano et al., 2009).

Las metodologías propias de estas áreas son diversas, y complementarias, presentando diferentes tipos y grados de ventajas y desventajas. Por ejemplo, en el caso del modelado matemático, este permite cuantificar la información disponible de un determinado fenómeno de una forma que es generalizable a otras situaciones o escalas, realizando predicciones que luego pueden ser contrastadas con nuevos experimentos, y permitiendo el diseño de intervenciones novedosas. Sin embargo, en muchos casos –sobre todo en los fenómenos sociales– los modelos matemáticos pueden acabar teniendo un dominio muy limitado a ciertos contextos simplificados, y no ser aplicables a entornos más realistas, o cotidianos (cuyo análisis es el objetivo verdadero de las ciencias sociales).

Otro ejemplo lo constituyen los experimentos comportamentales, que gozan de un alto grado de control, permitiendo establecer una relación causal entre las variables dependientes de interés y las variables independientes que son manipuladas. No obstante, los experimentos en ambientes muy controlados pueden carecer de validez ecológica, pudiendo argumentarse que las personas podrían no tomar las mismas decisiones en un laboratorio que en su vida cotidiana. A su vez, pueden ser muy costosos en términos de tiempo y dinero. Para mitigar estas desventajas, se ha vuelto muy común la utilización de plataformas online para la realización de estudios experimentales, así como también para estudios observacionales. Esto tiene varias ventajas, como la rapidez y el bajo costo para acceder a respuestas de muchas personas en diferentes lugares geográficos, facilitando enormemente la obtención de muestras con diferentes características (como, por ejemplo, muestras representativas). Sin embargo, también presenta desventajas, como sesgos de muestreo, mediciones más ruidosas, o respuestas de baja calidad (en contraposición con los experimentos tradicionales). Por ello, resulta fundamental intentar replicar los resultados en condiciones diversas (e.g., con muestras con diferentes características demográficas, incentivos, presiones temporales, etc.), así como preservar la rigurosidad y transparencia (e.g., realizando pre-registraciones de los análisis e hipótesis experimentales, y proveyendo libre acceso a los datos medidos y códigos de análisis y modelado).

En línea con las metodologías propias de estas áreas de investigación, en el Capítulo 2 de este trabajo se presenta un modelo matemático de un sesgo cognitivo ampliamente conocido y estudiado (Tversky & Kahneman, 1974). Se trata de un promedio pesado con pesos no lineales, con una forma exponencial decreciente, que cumple con ciertas propiedades en los límites de interés para el fenómeno estudiado. Este modelo predice la efectividad de un método propuesto

para mejorar la Sabiduría de las Multitudes. Para poner a prueba tanto el método como el modelo, se realizó una serie de experimentos comportamentales online (tres de ellos pre-registrados), que proveyeron evidencia convergente a favor de ambos. En el Capítulo 3 se analizaron estrategias microscópicas de interacción en grupos de discusión, tomando como punto de partida un estudio experimental previo (Navajas et al., 2018), y realizando tres estudios comportamentales online, el primero observacional, y los siguientes experimentales (y pre-registrados). A su vez, se desarrolló un método computacional basado en Procesamiento de Lenguaje Natural para analizar de forma automática la utilización de dichas estrategias microscópicas por parte de los grupos de discusión. Este trabajo contribuye a optimizar la Sabiduría de las Multitudes de una manera diferente, sugiriendo potenciales intervenciones y metodologías que podrían emplearse en contextos de discusiones online. Por último, en el Capítulo 4 se realizó un modelo computacional basado en agentes para analizar la emergencia de estados macroscópicos de consenso o bipolarización a partir de interacciones microscópicas con intercambios explícitos de argumentos en problemas no intelectivos. Este tipo de modelos permiten entender fenómenos sociales altamente complejos a partir de reglas simples de interacción microscópica.

#### **1.4. Estructura**

En suma, esta tesis plantea un puente entre la física estadística tradicional, la economía comportamental, y la psicología experimental, aplicadas a la profundización del entendimiento de las interacciones grupales, las variables que intervienen en la formación de consenso, y los

fundamentos de las decisiones democráticas, tanto fácticas como basadas en opinión. Se encuentra estructurada de la siguiente manera:

- En el capítulo 2 se desarrolla un método novedoso que utiliza el sesgo de anclaje para extremizar estimativos provistos por una multitud, con el objetivo de aumentar la diversidad de los mismos y mejorar la sabiduría de multitudes. Un modelo matemático del sesgo de anclaje, y cuatro estudios experimentales proveen evidencia convergente que respalda este método (Barrera et al, 2024).
- En el capítulo 3 se estudia un método previamente presentado por Navajas et al. (2018) con la intención de determinar qué estrategia utilizan grupos de personas que interactúan intentando llegar a un estimativo consensuado en problemas fácticos. Se desarrolla un método automático para detectar estas estrategias, y se convalida la hipótesis de que los grupos con mejor desempeño utilizan principalmente el Método de Fermi (Barrera et al, 2024, *en escritura*).
- En el capítulo 4 se expande un modelo de agentes previamente desarrollado por Balenzuela et al. (2015) para explicar los macroestados de consenso y bipolarización que pueden observarse en decisiones de opinión en nuestras sociedades, incorporando un intercambio explícito de argumentos, y evaluando el impacto de dos sesgos cognitivos muy conocidos (Barrera et al, 2020).
- En el capítulo 5 se presentan las conclusiones generales, las limitaciones y las perspectivas que se derivan de esta tesis.

Es importante destacar que los capítulos de la tesis están escritos y presentados de manera que puedan leerse de forma independiente, permitiendo abordar diferentes aspectos de las

interacciones grupales según los intereses específicos del lector. En cada capítulo se proporciona el contexto teórico necesario para comprender los análisis y resultados presentados, incluso si no se han leído los capítulos anteriores.

## Capítulo 2

### 2. Promover opiniones erróneas divergentes aumenta la sabiduría de las multitudes

#### 2.1. Resumen

La agregación de muchos juicios individuales genera estimaciones sorprendentemente exactas.

Este fenómeno, conocido como la "sabiduría de las multitudes", se ha demostrado en ámbitos diversos como la toma de decisiones médicas y las predicciones financieras. Muchas investigaciones previas identificaron dos factores clave que promueven este efecto: la exactitud de las evaluaciones individuales y la diversidad de opiniones. La mayoría de las estrategias disponibles para mejorar la sabiduría de las multitudes se han centrado en mejorar la exactitud individual, descuidando el potencial de aumentar la diversidad de opiniones. En este trabajo estudiamos un enfoque complementario para reducir el error colectivo al promover opiniones erróneas divergentes. Esta estrategia propone anclar a la mitad de la multitud a un valor pequeño, y a la otra mitad a un valor grande, antes de solicitar y promediar todas las estimaciones. Consistente con nuestro modelado matemático, cuatro experimentos demuestran que este método es efectivo tanto para tareas de estimación como de predicción. En general, esta investigación proporciona una herramienta práctica novedosa, y también contribuye a comprender el valor epistémico de la toma de decisiones colectiva.

## 2.2. Introducción

La agregación de muchas estimaciones individuales a menudo supera a los juicios individuales de expertos (De Condorcet, 1785; Galton, 1907). Este fenómeno, conocido popularmente como la "sabiduría de las multitudes" (Surowiecki, 2005), se ha aplicado a una amplia gama de problemas, como mejorar los diagnósticos médicos (Kurvers et al., 2016), prever eventos geopolíticos (Mellers et al., 2014), predecir los mercados financieros (Ray, 2006), realizar ingeniería inversa del olor de moléculas (Keller et al., 2017), y verificar hechos en las noticias (Allen, 2021), entre muchos otros. Dada su relevancia práctica, comprender las condiciones bajo las cuales las multitudes producen estimaciones exactas se ha convertido en un tema relevante en las ciencias psicológicas (Kameda, Toyokawa, y Tindale, 2022; Karachiwala y Pinkow, 2021; Navajas et al., 2018; Kao y Couzin, 2014).

Un impulsor importante de la exactitud colectiva es la diversidad de opiniones en la multitud (Hong y Page, 2004; Page, 2008; Becker, Porter y Centola, 2019; Shi et al., 2019; Jönsson, Hahn y Olsson, 2015). Una intuición simple subyace a esta afirmación: cuando las multitudes producen estimaciones diversas, es probable que algunos individuos subestimen la respuesta correcta, mientras que otros la sobreestimen. Por lo tanto, cuanto más diversa sea la multitud, mayor será la probabilidad de que los errores individuales se cancelen en el proceso de agregación. Más formalmente, el *Teorema de Predicción de Diversidad* (Page, 2007) establece que el error de la multitud ( $E$ ) se puede expresar como el error individual promedio ( $\varepsilon$ ) menos la diversidad predictiva de la multitud ( $\delta$ , también conocida como la varianza poblacional):

$$E = \varepsilon - \delta \quad [1]$$

Una implicancia de este teorema es que, en principio, la exactitud de la multitud podría aumentarse al reducir el error individual ( $\varepsilon$ ) o, de manera equivalente, al aumentar la diversidad predictiva ( $\delta$ ). Sin embargo, aunque estas dos estrategias son igualmente válidas en la teoría, la mayoría de los estudios disponibles que buscan aumentar la sabiduría de las multitudes se han centrado exclusivamente en reducir  $\varepsilon$ , descuidando el potencial de aumentar  $\delta$ .

Por ejemplo, estudios previos han propuesto agregar información de multitudes "selectas" compuestas por individuos que resulten ser más exactos en problemas de estimación (Mannes, Soll y Larrick, 2014). Otros estudios han demostrado que el error individual puede reducirse contrarrestando los sesgos individuales (Kao et al., 2018) o exponiendo a los individuos a información social (Jayles et al., 2017; Frey y Van de Rijt, 2021; Madirolas y de Polavieja, 2015; Lorenz et al., 2011). Una excepción notable a esta tendencia demostró que la exactitud colectiva puede aumentarse al mejorar la diversidad de procesos cognitivos (Keck y Tang, 2020), un constructo diferente al de diversidad predictiva. Sin embargo, incluso en ese caso, no está claro si el método realmente aumentó la diversidad de opiniones, redujo el error individual medio, o ambos. Algunos otros trabajos se han centrado en la diversidad, pero de manera menos directa. Por ejemplo, un artículo anterior realizó análisis secundarios para evaluar aumentos en la diversidad en concordancia con la teoría detrás de la Ecuación [1], pero sin manipular directamente la diversidad o la exactitud individual (Nobre & Fontanari, 2020). De manera similar, el fenómeno de la "multitud interna" (Vul & Pashler, 2008) –donde los experimentadores piden a cada participante que produzca más de una estimación–, se basa en la idea de aumentar la diversidad, pero dentro de los mismos individuos. Sin embargo, si y cómo este procedimiento aumenta la diversidad poblacional sigue siendo un interrogante abierto.

Demostrar que es posible disminuir el error colectivo simplemente aumentando la diversidad predictiva no es trivial, por varias razones. En primer lugar, demostraría que hay procesos que aumentan simultáneamente el error individual y la exactitud colectiva. En segundo lugar, desde un punto de vista práctico, les proporcionaría a los profesionales un enfoque novedoso para aumentar la exactitud de la estimación colectiva. En tercer lugar, dado que la sabiduría de las multitudes se ha interpretado previamente como evidencia empírica del valor epistémico de los juicios democráticos, esta supuesta disociación entre la exactitud individual y colectiva debería mitigar las preocupaciones sobre el aumento de votantes desinformados en elecciones recientes.

En conjunto, se converge hacia una posibilidad contraintuitiva, y quizás algo incómoda: si el error colectivo se reduce al aumentar la diversidad, esto implicaría que la sabiduría de las multitudes podría mejorarse persuadiendo a los individuos para que adopten opiniones erróneas divergentes. En este trabajo se presentan simulaciones teóricas y evidencia empírica para respaldar esta afirmación. Se expone un nuevo enfoque para aumentar la exactitud colectiva, potenciando la diversidad predictiva de una multitud, a expensas de reducir la exactitud individual, aún en casos en los que la verdad es completamente desconocida e inaccesible (incluso para los experimentadores).

### **2.3. Aumento de la Sabiduría de las Multitudes mediante Extremización**

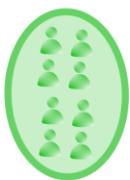
En este capítulo se propone fomentar estimaciones extremas –y, de esta forma, aumentar la diversidad–, mediante un sesgo cognitivo conocido como el “efecto de anclaje” (Tversky y Kahneman, 1974). El método planteado consiste en anclar a la mitad de la multitud a un valor

pequeño ("ancla baja"), y a la otra mitad a un valor grande ("ancla alta"), para luego promediar todas las estimaciones. Se hipotetiza que esta técnica conllevará a un aumento en la diversidad predictiva, el cual superará el aumento en el error individual medio, derivando así en una reducción del error colectivo. Utilizando un modelo matemático simple, primero se demuestra que este método mejora la exactitud colectiva en un amplio rango de parámetros. Luego, se corrobora el procedimiento empíricamente, mediante cuatro experimentos diferentes; y, finalmente, se muestra que este conduce, en efecto, a una reducción sustancial del error colectivo.

Como ejemplo, se considera el escenario en el que una persona necesita estimar una variable numérica que desconoce, por ejemplo, la altura de la Torre Eiffel. Basándose en el efecto estándar de la sabiduría de las multitudes, se podría obtener un valor aproximado al pedir a un gran número de individuos que proporcionen una estimación. Luego, para estimar dicha magnitud, se agregarían estos valores, por ejemplo, promediándolos (**Fig. 1A**). En este trabajo, se propone un enfoque alternativo que consiste en dividir a la multitud en dos mitades, y extremizar las opiniones en direcciones opuestas (**Fig. 1B**). Se sugiere hacerlo utilizando el efecto de anclaje: antes de estimar la variable relevante, se les pide a los individuos que primero consideren un valor extremadamente bajo o alto. En el ejemplo anterior, a la mitad de los individuos se les preguntaría si la altura de la Torre Eiffel es mayor o menor que 10 metros (ancla baja,  $A_L$ ) y a la otra mitad, si es mayor o menor que 1000 metros (ancla alta,  $A_H$ ). Despues de proporcionar una respuesta categórica a esta pregunta inicial, se les solicita a todos los individuos que proporcionen su mejor estimación. Mucha literatura previa ha evidenciado que estas estimaciones deberían estar sesgadas de forma consistente hacia los valores considerados

**A****PROCEDIMIENTO****Sabiduría de Multitudes**

¿Cuál es la altura  
de la Torre Eiffel?



$$\rightarrow \mu = 250 \text{ m}$$

**Sabiduría de Multitudes Extremizadas**

¿Es la altura de la  
Torre Eiffel mayor  
o menor a...



$$A_L = 10 \text{ m?} \rightarrow \mu_L = 100 \text{ m}$$

$$A_H = 1000 \text{ m?} \rightarrow \mu_H = 500 \text{ m}$$

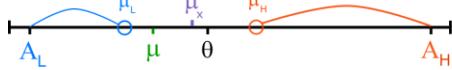
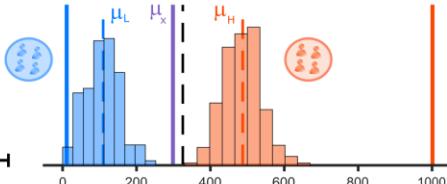
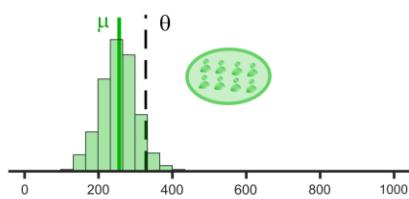
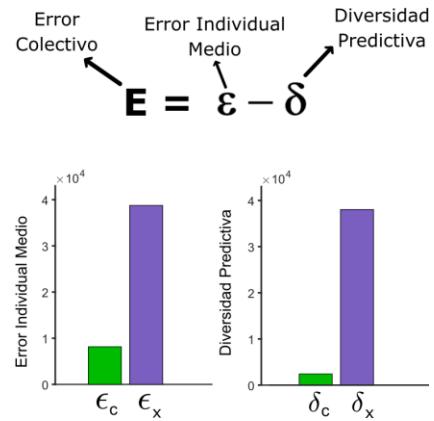
$$\frac{\mu_L + \mu_H}{2} = \mu_x = 300 \text{ m}$$

**B****MODELO**

$$\left\{ \begin{array}{l} \mu_L = w_L A_L + (1 - w_L)\mu \\ \mu_H = w_H A_H + (1 - w_H)\mu \end{array} \right.$$

**Efecto de Anclaje Extremo**

$$w_j(A_j) = w_0 e^{-\beta |A_j - \theta|}$$

**C****SIMULACIONES****D****PREDICCIONES**

**Fig. 1. Sabiduría de multitudes extremizadas.** (A) El método para estimar una variable a través de la sabiduría de las multitudes consiste en pedir a un grupo de individuos que estimen una cantidad dada, y promediar todas las respuestas. El método alternativo propuesto consiste en promover la adopción de opiniones divergentes dentro de una multitud dividiéndola en dos mitades, haciendo una pregunta de anclaje con un valor bajo o alto a cada mitad, y luego promediando todas las respuestas. (B) Modelo matemático del efecto de anclaje. La media anclada es un promedio ponderado del ancla y el valor de la sabiduría de las multitudes. El peso  $w$  depende de la diferencia entre el ancla y la respuesta correcta, reflejando una sensibilidad interna a la respuesta correcta. (C) Las simulaciones realizadas con el modelo propuesto muestran que es esperable que el método supere a la sabiduría de las multitudes. (D) El error individual medio y la diversidad predictiva en datos simulados muestran un aumento tanto en el error individual, como en la diversidad predictiva. Sin embargo, el aumento en la diversidad predictiva (panel derecho) es mayor que el aumento en el error individual (panel izquierdo), lo que resulta en una reducción general del error colectivo.

inicialmente (Furnham y Boo, 2011; Röseler et al., 2022). Dado que estas anclas son extremas en direcciones opuestas, este procedimiento debería radicalizar las estimaciones producidas por la multitud en su conjunto, llevando a un aumento en la diversidad predictiva. Por lo tanto, proponemos promediar todos los valores provistos (por ambas mitades de la multitud).

Aunque este procedimiento requiere predefinir dos valores extremos que se utilizarán como anclas, la estrategia no requiere conocer la respuesta correcta. Sin embargo, razonablemente, su exactitud dependerá de la elección específica de las anclas. Por lo tanto, para comprender mejor las condiciones bajo las cuales se espera que el enfoque propuesto aumente la exactitud colectiva, se desarrolla un modelo matemático simple del efecto de anclaje.

## 2.4. Modelo

Consideremos un conjunto de individuos que, al pedirles que estimen la variable  $\theta$ , producen una distribución de valores con una media  $\mu$ . El modelo asume que, cuando esos individuos están anclados a un valor bajo  $A_L$ , producen un conjunto de estimaciones con una media diferente:

$$\mu_L = w_L A_L + (1 - w_L) \mu, \quad [2]$$

donde  $0 \leq w_L \leq 1$  es un "índice de anclaje" que refleja el impacto del procedimiento de anclaje dado por el ancla  $A_L$ . De manera similar, una población anclada a un valor alto  $A_H$  produciría una distribución de estimaciones dada por

$$\mu_H = w_H A_H + (1 - w_H) \mu, \quad [3]$$

donde  $0 \leq w_H \leq 1$  es el "índice de anclaje" correspondiente, dado por el ancla  $A_H$ . Las ecs. [2] y [3] implican que la media anclada es un promedio ponderado del ancla y la media  $\mu$  de la distribución original de estimaciones.

Siguiendo una variedad de hallazgos empíricos que vinculan el efecto de anclaje con la plausibilidad del anclaje (Mussweiler y Strack, 2001; Wegener et al., 2001), proponemos que los pesos  $w$  se definan como

$$w_j = w_0 e^{-\beta |A_j - \theta|} \quad [4]$$

donde  $j$  indica si el peso corresponde al anclaje bajo o alto ( $w_L$  o  $w_H$ , respectivamente), y  $w_0$  es un parámetro que refleja el índice de anclaje cuando los individuos están anclados al valor correcto  $\theta$ . El parámetro  $\beta$  es una "temperatura inversa" que codifica la sensibilidad de los individuos a la distancia entre el ancla y  $\theta$ . Así, el valor de  $\beta$  modula la fuerza del "efecto de anclaje extremo" (Röseler et al., 2022).

En este trabajo, proponemos promediar las estimaciones de dos poblaciones de individuos, cada una de las cuales está anclada a un valor bajo o alto ( $A_L$  o  $A_H$ ). Suponiendo que ambas poblaciones tienen el mismo tamaño, podemos calcular el valor medio de la población en su conjunto como un promedio de los valores medios de cada una:

$$\mu_x = \frac{\mu_L + \mu_H}{2} \quad [5]$$

Las simulaciones (para un conjunto de parámetros empíricamente razonables, con  $\theta=324$ ,  $\mu=250$ ,  $A_L=10$ ,  $A_H=1000$ ,  $w_0=1$  y  $\beta=0.0017$ ) muestran que es esperable que este modelo produzca una reducción en el error colectivo (**Fig. 1C**). Este aumento en la exactitud colectiva va acompañado de un aumento en la diversidad, así como de una reducción en la exactitud

individual (**Fig. 1D**). La combinación de un aumento de diversidad acompañada por una reducción de exactitud individual es una característica subyacente que aparece bajo diferentes especificaciones del modelo (Fig. A1, Apéndice 1).

El método asume implícitamente que es posible seleccionar anclas de tal manera que el valor correcto será subestimado por el ancla baja y sobreestimado por el ancla alta. Sin embargo, cumplir con esta condición no es ni suficiente ni necesario. Por ejemplo, el método también debería funcionar cuando ambas anclas tienen el sesgo opuesto que las respuestas sin anclar. Por lo tanto, para entender las condiciones bajo las cuales se espera que funcione el método, realizamos un análisis detallado basado en el modelo previo.

Analíticamente, se observó que la variable clave que determina el éxito del enfoque es el valor medio de las anclas, definido como  $\bar{A} = \frac{A_L + A_H}{2}$ . Siguiendo un procedimiento matemático sencillo (para más detalles, consulte el Apéndice 1), encontramos que el rango de valores ( $\Delta$ ) de  $\bar{A}$  donde el método supera a la sabiduría de las multitudes es:

$$\Delta = \frac{4|\mu - \theta|}{w_L + w_H} \quad [6]$$

La expresión derivada en la Ec. [6] implica que el rango de valores donde el método supera a la sabiduría de las multitudes es siempre igual o mayor que dos veces el error colectivo. Esto se puede demostrar examinando dos escenarios extremos opuestos. Por un lado, si la sensibilidad  $\beta$  es pequeña (es decir, cuando el procedimiento de anclaje no depende de la distancia entre el ancla y la respuesta correcta), el rango de valores donde el método supera a la sabiduría de las multitudes converge, como mínimo, a dos veces el error colectivo (i.e., si  $\beta \rightarrow 0$ , entonces  $w_j \rightarrow w_0$ , y, por ende,  $\Delta \rightarrow \frac{2|\mu - \theta|}{w_0}$ ). En el caso opuesto, cuando la sensibilidad  $\beta$  es grande (es decir,

cuento el efecto de anclaje es más fuerte a medida que las anclas se acercan a la respuesta correcta), entonces este método siempre es superior a la sabiduría de las multitudes (i.e., if  $\beta \rightarrow \infty$ , then  $w_j \rightarrow 0$ , and thus  $\Delta \rightarrow \infty$ ). Para evaluar empíricamente la eficacia del método propuesto, realizamos cuatro experimentos.

## 2.5. Experimento 1: estimación de cantidades no acotadas

### 2.5.1 Métodos

#### *Participantes y Preguntas*

En el Experimento 1, N=120 participantes (48 mujeres, con una edad de  $37.2 \pm 11.6$  años, de los Estados Unidos, reclutados en línea a través de Amazon Mechanical Turk) proporcionaron estimaciones sobre 14 variables de conocimiento general (**Tabla S3**). Todas las variables eran positivas y no acotadas superiormente, al igual que el ejemplo utilizado en la **Figura 1** ("¿cuántos puentes hay en París?"). Los participantes tenían incentivos monetarios para estimar estas variables de la manera más exacta posible. Se informó a los participantes que su participación era completamente voluntaria y que podían retirarse en cualquier momento. Todos los datos eran completamente anónimos. El protocolo experimental fue aprobado por el comité de ética del Centro de Educación Médica e Investigaciones Clínicas Norberto Quirno (Buenos Aires, Argentina), protocolo 435, versión 5.

#### *Procedimiento*

Este experimento se desarrolló utilizando Psytoolkit (Stoet 2010; Stoet 2017). Un tercio de la muestra fue asignado aleatoriamente a una condición control donde simplemente estimaron

cada variable. Los otros dos tercios de los participantes fueron asignados aleatoriamente a la condición experimental donde, antes de estimar la cantidad, se les pidió que consideraran un valor extremadamente bajo o extremadamente alto (por ejemplo, ¿es el número de puentes en París mayor o menor que 349?). La mitad de los participantes anclados consideró un valor bajo, y la otra mitad consideró un valor alto (asignados aleatoriamente). Es crucial destacar que estos valores extremos no fueron elegidos manualmente por los experimentadores, sino que se establecieron automáticamente como los percentiles 5 y 95 de los valores empíricos observados en la condición de control (para este fin, los datos del primer tercio de la muestra se midieron antes que los datos restantes). En total, se obtuvieron datos de 41 participantes en la condición control, y 79 participantes en la condición experimental. Nuestra motivación para recolectar este número de participantes estuvo relacionada al tamaño máximo de multitud que es posible obtener para el procedimiento de resampling utilizado (ver “Análisis de Datos” abajo). Se buscó un tamaño máximo de multitud de por lo menos 30. En todos los casos, las preguntas se presentaron aleatoriamente. Todos los participantes tenían un máximo de 15 segundos para responder. Se pagó a los participantes una tarifa plana de 1.5 USD por su participación. La exactitud de la estimación se incentivó mediante el pago de un bono de 0.5 USD a los participantes en el 10% más exacto. Consulte la Tabla A2 en el Apéndice 1 para la lista completa de preguntas utilizadas en el Experimento 1.

## Análisis de Datos

Se descartaron los datos de participantes que completaron la encuesta en menos de tres minutos, y de aquellos que no lograron completarla. También se excluyó a los participantes con dos o más respuestas exactamente correctas (lo que es muy probable que refleje trampa).

Para comparar diferentes condiciones de manera equilibrada, se desarrolló una estrategia de resampling muy similar a *bootstrapping* (pero eligiendo muestras de distinto tamaño). Para cada tamaño de multitud, se seleccionó aleatoriamente (con reemplazo) un número fijo de individuos, y se estimaron el error colectivo, la diversidad predictiva, y el error individual medio, para ese tamaño de multitud y esa iteración. Repetimos ese procedimiento 1,000 veces para tamaños de multitud que variaban de 2 a distintos valores máximos en experimentos diferentes ( $n=32$  en el Experimento 1,  $n=50$  en el Experimento 2,  $n=100$  en el Experimento 3, y  $n=70$  en el Experimento 4). En todos los casos, esos valores permiten apreciar el comportamiento asintótico del error colectivo en función del tamaño de la multitud.

### 2.5.2 Resultados y Discusión

Al emplear un método sencillo de resampling mediante bootstrapping, se estimó el error colectivo de grupos de diferentes tamaños, tanto para la sabiduría de las multitudes clásica como para la sabiduría de las multitudes extremizadas (**Fig. 2A**). Observamos que el error colectivo medio de esta última fue siempre menor que el de la primera. Por ejemplo, el error colectivo de 34 individuos tomados al azar de la condición control fue sustancialmente mayor que el error colectivo de 34 individuos extremizados (test-t no apareado:  $t(998)=29.7$ ,  $p=2 \times 10^{-139}$ ; tamaño del efecto: delta de Cohen  $d = 1.8 \pm 0.1$ , 95% CL). Esta reducción en el error colectivo se debió a un

aumento en la diversidad predictiva (**Fig. 2B**, test-t no apareado:  $t(998)=54.8$ ,  $p<10^{-200}$ ; tamaño de efecto: delta de Cohen  $d = 3.3 \pm 0.1$ , 95% CL) que fue mayor que el aumento en el error individual medio (**Fig. 2B**, test-t no apareado:  $t(998)=24.2$ ,  $p=5\times 10^{-102}$  tamaño de efecto: delta de Cohen  $d = 1.4 \pm 0.1$ , 95% CL).

Una limitación del Experimento 1 es que las anclas se definieron después de recopilar los datos de la población no anclada. Dado que este procedimiento puede resultar inconveniente desde un punto de vista práctico, realizamos un segundo experimento pre-registrado donde las anclas fueron predefinidas y fijas para todas las preguntas ([https://aspredicted.org/RYC\\_4Y5](https://aspredicted.org/RYC_4Y5)).

## 2.6 Experimento 2: estimación de cantidades acotadas con anclas fijas

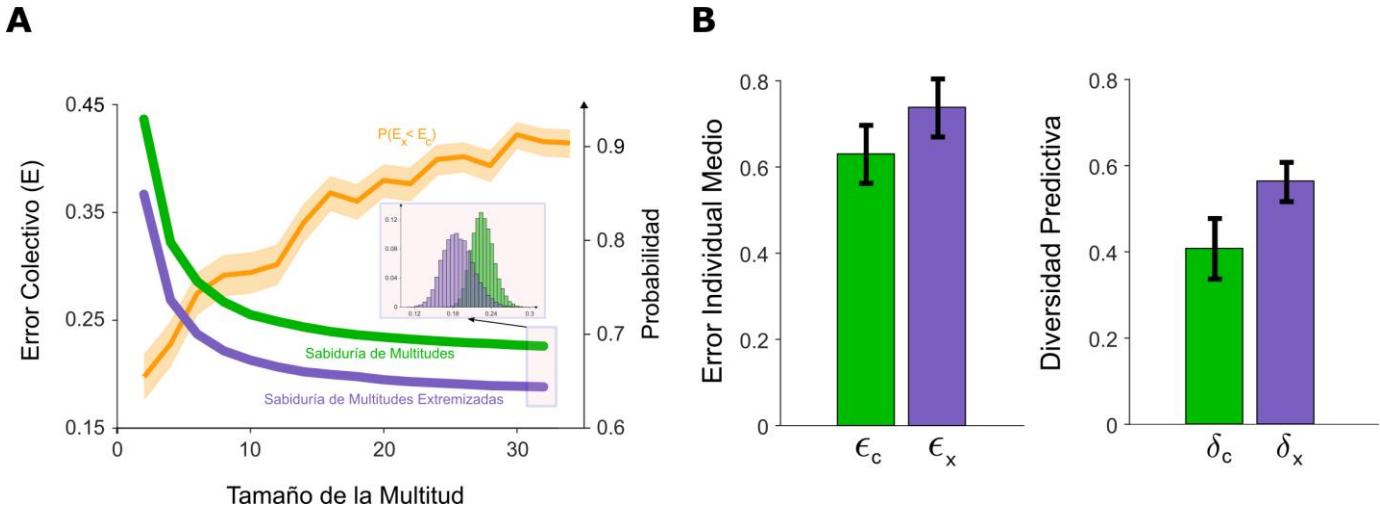
### 2.6.1 Métodos

#### *Participantes y Preguntas*

En el Experimento 2 se reclutaron  $N=396$  participantes en línea (235 mujeres, con una edad de  $27.9 \pm 8.8$  años) y se les hicieron 30 preguntas de conocimiento general, que en todos los casos implicaban la estimación de un porcentaje. Por lo tanto, todas las respuestas estaban acotadas en el rango [0,100] (por ejemplo, ¿qué porcentaje de la población de Argentina tiene menos de 15 años? Consulte la Tabla A3 en el Apéndice 1 para la lista completa de preguntas utilizadas en el Experimento 2). A diferencia del estudio anterior, aquí se utilizaron las mismas anclas para todas las preguntas, siempre establecidas en 5% (anclaje bajo  $A_L$ ) o 95% (anclaje alto  $A_H$ ). Los

## Experimento 1 (N=120)

Tipo de Pregunta: ¿cuál es la distancia en millas entre Atenas y Roma?



**Fig. 2. Resultados empíricos para el Experimento 1. (A)** Error colectivo en función del tamaño de la multitud, para la sabiduría de las multitudes clásica (verde) y para las multitudes extremizadas (violeta). El error estándar de cada curva está dentro del ancho de las líneas. El recuadro muestra la distribución de valores del método de resampling para el tamaño de multitud más grande ( $N=34$ ). La línea naranja representa la probabilidad de que una muestra de la distribución de errores de la multitud extremizada (distribución violeta,  $E_x$ ) sea menor que una muestra de la distribución de errores de la multitud no extremizada (distribución verde,  $E_c$ ). **(B)** Error individual medio y diversidad predictiva tanto para la sabiduría de las multitudes clásica (verde) como para la sabiduría de las multitudes extremizadas (violeta). Las barras de error constituyen la desviación estándar de las medias.

participantes fueron reclutados en línea y residían en Argentina en el momento del experimento. Todas las preguntas provenían de diversas encuestas representativas realizadas por entidades oficiales. El contenido de las preguntas estaba explícitamente relacionado con Argentina, y abarcaba desde aspectos demográficos (por ejemplo, ¿qué porcentaje de la población mayor de

20 años tiene sobrepeso u obesidad?) hasta percepciones personales sobre el país (por ejemplo, ¿qué porcentaje de la población cree que el aborto es moralmente inaceptable?). Cuando fue necesario, se referenció el año en que se llevó a cabo la encuesta correspondiente.

### *Procedimiento*

Este experimento también se desarrolló utilizando Psytoolkit (Stoet 2010; Stoet 2017). Para reducir la longitud de la encuesta, se dividieron las preguntas en dos conjuntos de 15 preguntas. Cada participante fue asignado aleatoriamente a uno de estos conjuntos de preguntas. Además, siguiendo un procedimiento muy similar al experimento anterior, se dividió la muestra aleatoriamente en dos grupos. A los participantes en la condición control se les pidió que estimaran directamente la respuesta a las 15 preguntas (por ejemplo, ¿qué porcentaje de la población de Argentina es menor de 18 años?). Los participantes en la condición experimental respondieron las mismas preguntas, pero después de una pregunta de "anclaje" (¿creés que el porcentaje de la población de Argentina que es menor de 18 años es superior o inferior al 95%?). La asignación a la condición de ancla baja o ancla alta fue aleatoria en todas las preguntas. En total, se reunieron datos de 119 participantes en el Grupo 1, y 277 participantes en el Grupo 2. En todos los casos, las preguntas se presentaron en orden aleatorio (dentro de cada conjunto de 15 preguntas). En este caso, se estimó el número de participantes por condición requerido a partir de análisis de potencia estadística derivado del experimento previo (siguiendo la pre-registración, [https://aspredicted.org/RYC\\_4Y5](https://aspredicted.org/RYC_4Y5)). Todos los participantes tuvieron 20 segundos para responder cada pregunta. La estimación no fue incentivada económicamente por exactitud.

## *Análisis de Datos*

Para este experimento se siguieron los mismos criterios de exclusión que para el anterior. En este caso, los criterios de exclusión fueron pre-registrados. Para el Experimento 2 se siguió el mismo procedimiento de resampling mediante bootstrapping que para el experimento anterior.

### **2.6.2 Resultados y Discusión**

Se observaron resultados muy similares a los del Experimento 1 (**Fig. 3A y 3B**). El error colectivo fue menor para la multitud extremizada (tamaño de la multitud más grande: N=50, test-t no apareado:  $t(998)=19.1$ ,  $p=2 \times 10^{-69}$ ; tamaño de efecto: delta de Cohen  $d = 1.2 \pm 0.1$ , 95% CL). Esto fue acompañado por un aumento en el error individual medio test-t no apareado:  $t(198)=121.1$ ,  $p<10^{-200}$ ; tamaño de efecto: delta de Cohen  $d = 7.6 \pm 0.3$ , 95% CL) así como un aumento en la diversidad predictiva (test-t no apareado:  $t(198)=162.5$ ,  $p<10^{-200}$ ; tamaño de efecto: delta de Cohen  $d = 10.3 \pm 0.3$ , 95% CL).

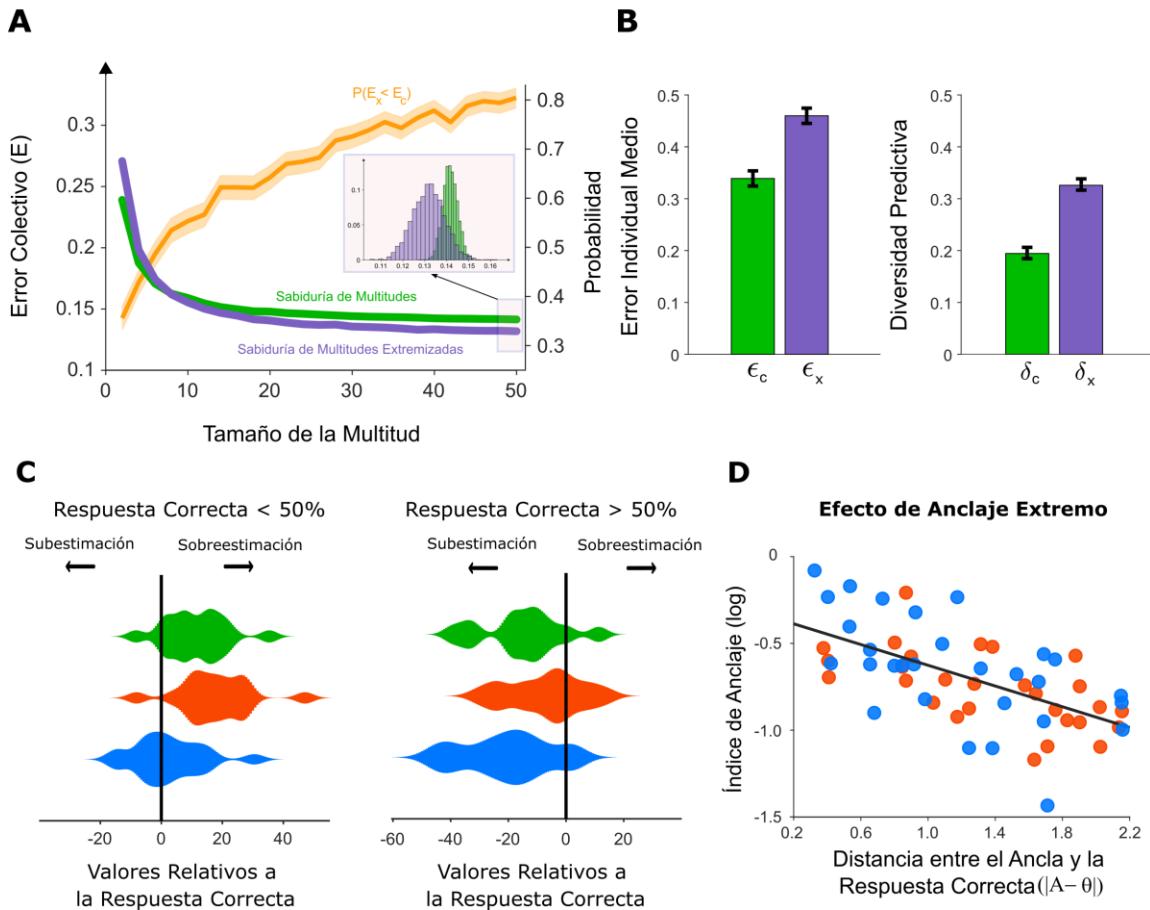
Dado que este experimento utilizó anclas fijas para todas las preguntas, esto permitió corroborar un elemento clave del modelo: el efecto de anclaje extremo (Ec. [4]). Esto se realizó por medio de dos análisis separados. En primer lugar, se examinaron los sesgos asociados a cada condición experimental. Se debe notar que, si el anclaje es sensible a la distancia del ancla con la respuesta correcta, entonces la efectividad del procedimiento debería ser mayor cuando la respuesta correcta es más cercana a un ancla específica. Por ejemplo, deberíamos ver que los participantes que consideran un valor bajo (5%) deberían estar más atraídos hacia el ancla cuando la respuesta correcta es baja (por debajo del 50%) en comparación con cuando la respuesta correcta es alta (por encima del 50%). Consistente con esta idea, cuando la respuesta

correcta se encontraba por debajo del 50% (17 preguntas), observamos que la población que consideraba el ancla baja proporcionaba una distribución de estimaciones similar a la respuesta correcta (test-t pareado,  $t(16)=0.95$ ,  $p=.36$ , tamaño de efecto: delta de Cohen  $d = 0.23 \pm 0.67$ , 95% CL), y un análisis del Factor de Bayes provee apoyo moderado a la hipótesis nula (factor de Bayes = 2.7). A su vez, tanto la población no anclada (test-t pareado,  $t(16)=4.16$ ,  $p=3\times 10^{-4}$ , tamaño de efecto: delta de Cohen  $d = 1.11 \pm 0.72$ , 95% CL) como la población anclada a un valor alto (test-t pareado,  $t(16)=5.99$ ,  $p=2\times 10^{-5}$ , tamaño de efecto: delta de Cohen  $d = 1.45 \pm 0.76$ , 95% CL) proporcionaron una distribución de estimaciones que sobreestimaba significativamente la respuesta correcta (**Fig. 3C**). Por el contrario, cuando la respuesta correcta estaba por encima del 50% (13 preguntas), observamos el patrón opuesto: la población que consideraba el ancla alta proporcionaba una distribución de estimaciones similar a la respuesta correcta (test-t pareado,  $t(12)=2.01$ ,  $p=0.07$ , tamaño de efecto: delta de Cohen  $d = 0.56 \pm 0.78$ , 95% CL), y el análisis de Factor de Bayes sugiere evidencia muy débil a favor de la hipótesis alternativa (factor de Bayes = 0.77). Crucialmente, tanto la población no anclada (test-t pareado,  $t(16)=3.90$ ,  $p=0.002$ , tamaño del efecto: delta de Cohen  $d = 1.08 \pm 0.82$ , 95% CL) como la población anclada a un valor bajo (test-t pareado,  $t(16)=4.90$ ,  $p=4\times 10^{-4}$ , tamaño del efecto: delta de Cohen  $d = 1.36 \pm 0.85$ , 95% CL) proporcionaron una distribución de estimaciones que subestimaba significativamente la respuesta correcta.

En segundo lugar, se examinó el efecto de anclaje extremo directamente, estudiando la asociación entre el impacto del procedimiento de anclaje y la distancia entre el ancla y la respuesta correcta (Ec. [4]). Se estimó el impacto del efecto de anclaje calculando el "índice de anclaje", que es similar al que se observa en literatura previa (Jacowitz & Kahneman, 1995). El

## Experimento 2 (N=396)

Tipo de Pregunta: ¿qué porcentaje de la población de Argentina es menor de 15 años?



**Fig. 3. Resultados empíricos para el Experimento 2.** (A) Error colectivo en función del tamaño de la multitud, tanto para la sabiduría de las multitudes clásica (verde) como para las multitudes extremizadas (violeta). El error estándar de las curvas está dentro del ancho de las líneas. El recuadro muestra la distribución de valores del método de resampling para el tamaño de multitud más grande ( $N=50$ ). La línea naranja representa la probabilidad de que una muestra de la distribución de errores de la multitud extremizada (distribución violeta,  $E_x$ ) sea menor que una muestra de la distribución de errores de la multitud no extremizada (distribución verde,  $E_c$ ). (B) Error individual medio y diversidad predictiva tanto para la sabiduría de las multitudes clásica (verde) como para las multitudes extremizadas (violeta). Las barras de error muestran la desviación estándar de las medias. (C) Distribuciones de valores correspondientes a la diferencia entre las respuestas medias y la respuesta correcta para cada pregunta, tanto para la sabiduría de las multitudes clásica (verde), la multitud extremizada usando un ancla alta (rojo) y la multitud extremizada usando un ancla baja (azul). Se separan los casos donde la respuesta correcta está por encima del 50% (panel izquierdo) y donde está por debajo del 50% (panel derecho). La

Línea negra representa el caso donde el valor medio es igual a la respuesta correcta. (**D**) Índice de anclaje empírico ( $w$ ) para cada pregunta. Los puntos azules muestran estimaciones usando anclas bajas y los puntos rojos muestran lo mismo con anclas altas. El eje horizontal representa la distancia entre el ancla correspondiente y la respuesta correcta (unidades logarítmicas en el índice de anclaje), y la línea negra muestra el mejor ajuste lineal de los datos.

índice (ver Apéndice 1 para más detalles) toma un valor de 0 cuando el procedimiento de anclaje no produce ningún efecto en las estimaciones, y un valor de 1 cuando las estimaciones son, en promedio, iguales al ancla. Consistente con la existencia del efecto de anclaje extremo (Roseler et al., 2022), se observó una correlación negativa significativa entre el índice de anclaje y la distancia entre el ancla y la respuesta correcta (coeficiente de correlación de Pearson,  $r=-0.56$ ,  $p=3 \times 10^{-6}$ ). Esta observación empírica respalda el modelo propuesto del efecto de anclaje utilizado en las Ecuaciones 2-4.

Para evaluar la forma funcional del modelo propuesto, se lo comparó con otras dos funciones monótonas decrecientes: una función lineal, y una función hiperbólica general:

$$w = \frac{w_0}{1 + \beta |A - \theta|} \quad [7]$$

similar a la que se utiliza en la literatura de descuento temporal (Ruggeri et al., 2022). Se observó que el modelo exponencial deriva en un mejor ajuste a los datos, evaluado con los Criterios de Información de Akaike y Bayesiano (modelo lineal:  $\Delta AIC = 1.9$ ,  $\Delta BIC = 1.9$ ; modelo hiperbólico:  $\Delta AIC = 0.59$ ,  $\Delta BIC = 0.59$ ).

Otra suposición del modelo de anclaje propuesto es que las anclas bajas y altas derivan en los mismos índices de anclaje. Dado que esta característica del modelo podría ser una sobre-simplificación para problemas específicos (por ejemplo, poblaciones de ciudades), se comparó la

bondad del ajuste obtenido con un único índice de anclaje por pregunta, con el correspondiente a usar índices diferentes para anclas bajas y altas. La comparación de modelos indicó que tener un único índice por pregunta deriva en mejores ajustes que el modelo con índices asimétricos ( $\Delta\text{AIC} = 1.6$ ,  $\Delta\text{BIC} = 5.7$ ).

Por último, este experimento permitió evaluar si el método propuesto es robusto al uso de diferentes procedimientos de agregación como el promediado simple, la mediana, y un promedio pesado por desempeño (Mannes et al., 2014; Collins et al, 2023). Se encontró evidencia de que, independientemente del método de agregación específico, las multitudes extremizadas por medio del sesgo de anclaje fueron siempre más exactas que las no extremizadas (Fig. A3; para más detalles, ver el Apéndice 1).

## 2.7 Experimento 3: predicción

Subsecuentemente, se plantea la duda de si el método propuesto puede ser útil para superar la sabiduría de las multitudes clásica en tareas de predicción o pronóstico, es decir, en dominios donde la respuesta correcta es desconocida e inaccesible en el momento del experimento. Para responder a esta pregunta, se realizó un tercer experimento pre-registrado que tuvo lugar en medio de la crisis de la COVID-19 ([https://aspredicted.org/HZC\\_PTH](https://aspredicted.org/HZC_PTH)).

### 2.7.1 Métodos

#### *Participantes y Preguntas*

Se reclutaron N=620 participantes (312 mujeres, con una edad de  $46.1 \pm 15.7$  años) de Estados Unidos, y se les solicitó que estimaran el número total de casos y muertes por COVID-19 que

ocurrirían en ese país en la semana siguiente (del 27 de julio al 2 de agosto de 2020). Por lo tanto, las respuestas eran positivas (no acotadas), de forma similar a los experimentos típicos de sabiduría de las multitudes (y al primer experimento de este capítulo), pero relacionadas con cantidades desconocidas en el momento del experimento. Los participantes fueron reclutados en línea utilizando Prolific (<https://www.prolific.co/>), y residían en Estados Unidos al momento del experimento. Los participantes tenían incentivos económicos para estimar estas variables con la mayor exactitud posible. Las anclas se seleccionaron como valores extremos basados en datos históricos (en la Tabla S4 en el Apéndice 1 se presentan las preguntas utilizadas en el Experimento 3).

#### *Procedimiento*

En el Experimento 3, desarrollado utilizando Survey Monkey (<https://www.surveymonkey.com/>), se hicieron dos preguntas relacionadas con la pandemia de COVID-19: se les pidió a los participantes que pronosticaran el número de muertes y casos de COVID-19 en la semana siguiente al experimento. Las anclas se seleccionaron como valores extremos basados en datos históricos, es decir, dos órdenes de magnitud menos o más que el número de casos y muertes por COVID-19 informados en las dos semanas anteriores al inicio del experimento. Las respuestas correctas a esas preguntas eran desconocidas en el momento del experimento. Dado que todos los participantes respondieron ambas preguntas, hubo un total de seis condiciones. En la Condición 1, se utilizó un ancla baja en las muertes por COVID-19. Se les preguntó a los participantes si creían que habría más o menos de 40 nuevas muertes en la semana siguiente al experimento, y luego se les pidió que pronosticaran el número de nuevas muertes. En la siguiente pantalla, los participantes también pronosticaron el número de nuevos casos de COVID-19 en la

semana siguiente al experimento. La Condición 2 fue la misma que la Condición 1, pero con un ancla alta (400,000) en las muertes por COVID-19. En la Condición 3 no se utilizaron anclas (lo que sirvió como control para las Condiciones 1 y 2). Primero se preguntó cuántas nuevas muertes pronosticaban en la semana siguiente. En la siguiente pantalla, los participantes pronosticaron el número de nuevos casos de COVID-19 en la misma semana. La Condición 4 fue análoga a la Condición 1, pero cambiando el orden de las preguntas y estableciendo un ancla en los casos, en lugar de en las muertes. Primero, se consultó a los participantes si creían que habría más o menos de 8,000 nuevos casos en la semana siguiente, y luego se les pidió que estimaran el número de nuevos casos. En la siguiente pantalla, los participantes pronosticaron el número de nuevas muertes por COVID-19 en la misma semana. La Condición 5 fue la misma que la Condición 4, pero con un ancla alta (8,000,000) en los casos de COVID-19. En la Condición 6, no se hicieron preguntas de anclaje, y se pidió a los participantes que pronosticaran directamente el número de nuevos casos, y luego el número de nuevas muertes, en la semana siguiente al experimento. Esta condición fue análoga a la Condición 3, pero cambiando el orden de las preguntas, por lo que sirvió como control para las Condiciones 4 y 5.

Se obtuvieron datos de 117 participantes en la Condición 1, 105 participantes en la Condición 2, 97 participantes en la Condición 3, 92 participantes en la Condición 4, 97 participantes en la Condición 5 y 112 participantes en la Condición 6. La muestra total de 600 participantes, obtenida a través de Prolific (<https://www.prolific.co/>), fue representativa de la población estadounidense en cuanto a edad, género, y etnia. En este caso, se buscó contar con 100 participantes por condición, para disponer de un tamaño máximo de multitud de 100 para el análisis de resampling (según fue pre-registrado en [https://aspredicted.org/HZC\\_PTH](https://aspredicted.org/HZC_PTH)). Todos los participantes

recibieron una tarifa plana de participación de 1.0 USD y, para incentivar la exactitud en el pronóstico, se les pagó un bono de 2.0 USD a los participantes dentro del 10% más exacto. No hubo límite de tiempo para responder estas preguntas.

### *Análisis de Datos*

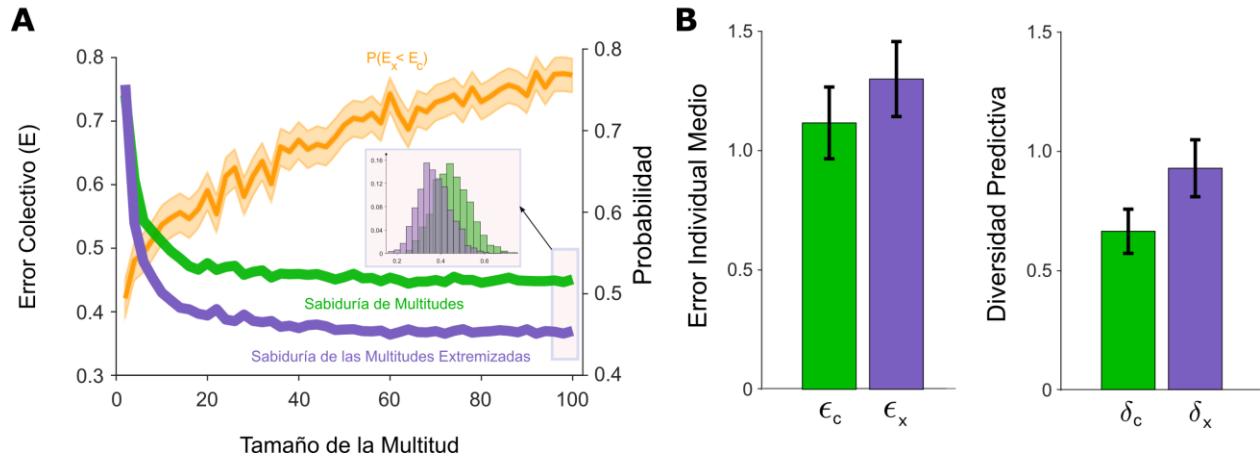
Para el Experimento 3, dado que consistió en preguntas de pronóstico, y que cada participante completó la encuesta, no fue necesario excluir a ninguno de ellos. Todos respondieron ambas preguntas de pronóstico. Seguimos el mismo procedimiento de resampling, similar a bootstrapping, que se empleó en el análisis de los experimentos anteriores.

## **2.7.2 Resultados y Discusión**

Nuevamente, como en los Experimentos 1 y 2, observamos que el error colectivo fue menor para la multitud extremizada en todos los tamaños de grupo, en comparación con la sabiduría de las multitudes clásica (**Fig. 4A**). Encontramos que la disminución del error colectivo (para el tamaño de grupo más grande N=100, test-t no apareado:  $t(998)=16.4$ ,  $p=7\times 10^{-54}$ ; tamaño del efecto: delta de Cohen  $d = 1.04 \pm 0.09$ , 95% CL) se debió a un aumento en la diversidad predictiva (para el tamaño de grupo N=100, test-t no apareado:  $t(998)=37.6$ ,  $p=7\times 10^{-193}$ ; tamaño del efecto: delta de Cohen  $d = 2.5 \pm 0.1$ , 95% CL). Este aumento en la diversidad predictiva fue a su vez mayor que el aumento en el error individual medio (**Fig. 4B**, tamaño de grupo N=100, test-t no apareado:  $t(998)=18.1$ ,  $p=1\times 10^{-64}$ ; tamaño del efecto: delta de Cohen  $d = 1.2 \pm 0.1$ , 95% CL). De esta manera, este estudio demuestra que la estrategia presentada también es útil para una tarea de predicción.

## Experimento 3 (N=620)

Tipo de Pregunta: ¿cuántas muertes por COVID-19 ocurrirán en USA la próxima semana?



**Fig. 4. Resultados Empíricos para el Experimento 3.** (A) Error colectivo en función del tamaño de la multitud para la sabiduría de las multitudes clásica (verde), y las multitudes extremizadas (violeta). El error estándar de las curvas está dentro del ancho de las líneas. El recuadro muestra la distribución de valores del método de resampling para el tamaño de grupo más grande (N=100). La línea naranja representa la probabilidad de que una muestra de la distribución de errores de la multitud extremizada (distribución violeta,  $E_x$ ) sea menor que una muestra de la distribución de errores de la multitud no extremizada (distribución verde,  $E_c$ ). (B) Error individual medio y diversidad predictiva tanto para la sabiduría de las multitudes clásica (verde) como para la multitud extremizada (violeta). Las barras de error corresponden a la desviación estándar de las medias.

## **2.8 Experimento 4: anclas auto-generadas**

Una limitación de los Experimentos 1-3 es que, en todos los casos, los participantes tenían acceso a información adicional (ya sea porque las anclas fueron predefinidas por los experimentadores o basadas en las respuestas dadas por otros participantes). El Experimento 4 abordó esa limitación utilizando anclas generadas por los propios participantes. El experimento fue pre-registrado ([https://aspredicted.org/YQ4\\_LYS](https://aspredicted.org/YQ4_LYS)) y utilizó preguntas similares a las utilizadas en el Experimento 1. La diferencia clave es que se pidió a los participantes que generaran sus propias anclas.

### **2.8.1 Métodos**

#### *Participantes y Preguntas*

Se reclutaron N= 226 participantes (110 mujeres, con una edad de  $37.2 \pm 12.1$  años, residentes en Argentina al momento del experimento, reclutados en línea) y se les hicieron 10 preguntas de conocimiento general que implicaban la estimación de la altura de un edificio (por ejemplo, ¿cuál es la altura de la Torre Eiffel?). A diferencia de todos los experimentos anteriores, las anclas fueron generadas por los propios participantes, lo que significa que proporcionaron sus propias anclas respectivas (consultar la Tabla A5 en el Apéndice 1 para ver la lista completa de preguntas).

#### *Procedimiento*

Este experimento fue desarrollado utilizando Psytoolkit (Stoet 2010; Stoet 2017). Un tercio de la muestra fue asignado aleatoriamente a una condición de control, donde simplemente estimaron una variable. Después de que los participantes completaron las diez preguntas, se les pidió que proporcionaran nuevas estimaciones diferentes para cada una de ellas. Este procedimiento nos

permittió comparar si la generación propia de anclas en las condiciones de tratamiento era diferente o no al fenómeno de “bootstrapping dialéctico” (Herzog & Hertwig, 2009).

Los otros dos tercios de los participantes fueron asignados aleatoriamente a la condición experimental donde, antes de comenzar el cuestionario principal, se les pidió que estimaran la altura del edificio más alto de la Tierra y la altura del edificio más bajo de la Tierra. Los valores que proporcionaron se utilizaron como las anclas altas y bajas, respectivamente, para todas las preguntas de estimación. Se agregó ruido uniforme del 10% con media cero a estas estimaciones preliminares, de modo que cada pregunta con la misma ancla de base tuviera un valor ligeramente diferente. Así, antes de estimar la cantidad en cuestión, se les pidió que consideraran un valor extremadamente bajo o un valor extremadamente alto (por ejemplo, ¿es la altura de la Torre Eiffel mayor o menor que 1000 metros?). Para cada pregunta, la mitad de los participantes anclados consideró un valor bajo, y la otra mitad consideró un valor alto (asignado aleatoriamente). En total, recopilamos datos de 75 participantes en la condición de control y 151 participantes en la condición experimental. El tamaño de muestra para cada condición se determinó mediante análisis de potencia estadística (según la pre-registración, para más detalles ver [https://aspredicted.org/YQ4\\_LYS](https://aspredicted.org/YQ4_LYS)). En todos los casos, las preguntas se presentaron en orden aleatorio. Todos los participantes tuvieron un máximo de 25 segundos para responder. La estimación no fue incentivada por exactitud.

### *Análisis de Datos*

Se descartaron los datos de los participantes que completaron la encuesta en menos de dos minutos, o aquellos que no lograron completar la mitad o más de las preguntas en la encuesta. También se excluyeron a los participantes con dos o más respuestas exactamente correctas (lo

que probablemente refleja trampas). Estos fueron los mismos criterios de exclusión que para los Experimentos 1 y 2. Estos criterios de exclusión fueron pre-registrados. Para el Experimento 4 seguimos el mismo procedimiento de resampleo que se indicó para los experimentos anteriores.

## 2.8.2 Resultados y Discusión

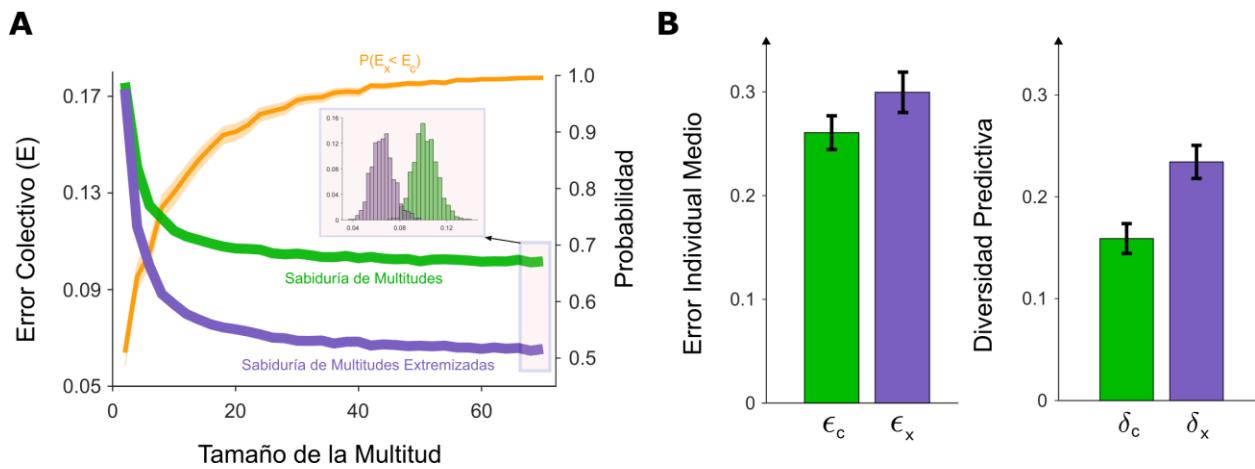
Al igual que en los experimentos anteriores, se observó que el error colectivo fue menor para la multitud extremizada para todos los tamaños de grupo en comparación con la sabiduría de las multitudes no extremizada (Fig. 5A). Encontramos que la disminución en el error colectivo (específicamente, para el tamaño de multitud más grande N=70, test-t no apareado:  $t(998)=58.4$ ,  $p<10^{-200}$ ; tamaño de efecto: delta de Cohen  $d = 3.7 \pm 0.1$ , 95% CL) fue debido a un incremento en diversidad predictiva (tamaño de multitud N=70, test-t no apareado:  $t(998)=78.5$ ,  $p<10^{-200}$ ; tamaño de efecto: delta de Cohen  $d = 4.9 \pm 0.2$ , 95% CL). Este incremento en diversidad predictiva fue a su vez mayor que el incremento en error individual medio (Fig. 5B, tamaño de multitud N=70, test-t no apareado:  $t(998)= 33.9$ ,  $p=2\times 10^{-168}$ ; tamaño de efecto: delta de Cohen  $d = 2.2 \pm 0.1$ , 95% CL). En general, este estudio demuestra que la estrategia presentada también es útil al emplear anclas generadas por los propios participantes. Esto significa que el problema de elegir anclas apropiadas puede resolverse fácilmente al pedir a los participantes que proporcionen sus propias anclas.

A continuación, se estudió la relación entre este procedimiento y el efecto de la sabiduría de la multitud interna. Dado que los participantes generaron sus propias anclas, se razona que este efecto podría, en principio, explicar parcialmente la eficacia del método. Sin embargo, se logró descartar esta posibilidad, ya que la sabiduría de las multitudes extremizadas produjo

estimaciones colectivas mucho más exactas que la sabiduría de la multitud interna (para más detalles, remitirse al apartado "Anclas Auto-Generadas y la Sabiduría de la Multitud Interna" en el Apéndice 1).

## Experimento 4 (N=226)

Tipo de Pregunta: ¿Cuál es la Altura de la Torre Eiffel? (con anclas auto-generadas)



**Fig. 5. Resultados Empíricos para el Experimento 4.** (A) Error colectivo en función del tamaño de la multitud para la sabiduría de las multitudes clásica (verde), y las multitudes extremizadas (violeta). El error estándar de las curvas está dentro del ancho de las líneas. El recuadro muestra la distribución de valores del método de resampling para el tamaño de grupo más grande ( $N=70$ ). La línea naranja representa la probabilidad de que una muestra de la distribución de errores de la multitud extremizada (distribución violeta,  $E_x$ ) sea menor que una muestra de la distribución de errores de la multitud no extremizada (distribución verde,  $E_c$ ). (B) Error individual medio y diversidad predictiva tanto para la sabiduría de las multitudes clásica (verde) como para la multitud extremizada (violeta). Las barras de error corresponden a la desviación estándar de las medias.

## 2.9 Discusión General

En este capítulo se presentó una estrategia novedosa para mejorar la sabiduría de las multitudes (Surowiecki, 2005), con implicancias tanto prácticas como teóricas. Mediante el empleo del efecto de anclaje (Tversky y Kahneman, 1974), se demostró que es posible reducir el error colectivo aumentando la diversidad predictiva (Page, 2007). En la literatura previa, los métodos para aumentar la sabiduría de las multitudes a menudo involucraban estrategias para reducir el error individual (Madirolas y de Polavieja, 2015; Mannes, Soll y Larrick, 2014). Sin embargo, aunque el análisis teórico sugería que este objetivo también se podría lograr aumentando la diversidad predictiva (como se ve en la Ecuación [1]), esa posibilidad fue muy poco explorada hasta ahora. En este capítulo se estudió este enfoque en profundidad, de manera tanto teórica como empírica: primero, desarrollando un modelo matemático del efecto de anclaje y, segundo, realizando cuatro experimentos comportamentales.

En todos los experimentos, independientemente de las diferencias en el tamaño de la muestra, el país en el que se implementó, el uso de cantidades acotadas o no acotadas, y si la tarea involucraba estimación o pronóstico, se observaron resultados muy similares. Las multitudes extremizadas siempre producían estimaciones con un error colectivo más bajo, y esto siempre iba acompañado de un aumento tanto en el error individual medio, como en la diversidad predictiva. Esto demuestra que es posible aumentar la inteligencia colectiva, al tiempo que se reduce la exactitud individual, un enfoque que hasta ahora no se había probado empíricamente. Por lo tanto, esto debería inspirar futuras investigaciones destinadas a aumentar la sabiduría de las multitudes mediante una estrategia similar.

Una limitación del método propuesto es que la selección de anclas apropiadas podría resultar potencialmente difícil (por ejemplo, en problemas de pronóstico). Sin embargo, tanto el análisis del modelo presentado, como los resultados empíricos encontrados, sugieren que el rango de valores donde este método mejora la sabiduría de las multitudes es amplio (Ecuación 4). También se exhibió que seleccionar anclas de manera adecuada es empíricamente factible en cuatro configuraciones experimentales muy diferentes. En el Experimento 1, las anclas se eligieron en base a valores extremadamente altos y bajos (percentiles 5 y 95) de la distribución no anclada. Sin embargo, este procedimiento tiene la desventaja de compartir información entre diferentes grupos de participantes, lo que podría explicar en parte los cambios en el rendimiento colectivo (por ejemplo, Navajas et al., 2018; Becker et al., 2017; Becker et al., 2021). En el Experimento 2, estudiamos la estimación de porcentajes manteniendo las anclas fijas en 5% y 95%. Con este experimento se demostró que el método funciona incluso en una condición donde las anclas fueron predefinidas para todas las preguntas. Esta configuración también nos permitió obtener sustento empírico para el modelo de anclaje propuesto. En el Experimento 3, se varió el enfoque hacia problemas de pronóstico. Las anclas se definieron de antemano, utilizando datos históricos. Dado que la respuesta correcta era desconocida para los experimentadores antes de realizar el estudio, esta configuración mostró que, incluso en problemas de pronóstico altamente inciertos, es factible seleccionar anclas para que el método funcione. Sin embargo, una limitación de los Experimentos 1-3 es que, en todos los casos, los participantes tenían acceso a información adicional (ya sea porque las anclas fueron predefinidas por los experimentadores, basadas en las respuestas dadas por otros participantes, o en series históricas). El Experimento 4 abordó esa limitación utilizando anclas generadas por los propios participantes. Esto demuestra que este

método es útil incluso en condiciones donde las anclas no son determinadas por los experimentadores.

El primer estudio sobre la sabiduría de las multitudes (Galton, 1907) fue considerado como una demostración empírica de que los principios de agregación democrática son confiables y eficientes. Esto fue contrario a la intuición en ese momento, ya que mostraba que individuos erróneos podían tomar buenas decisiones colectivas. Hoy en día, cuando las opiniones políticas tienden a polarizarse, estos resultados parecen sugerir que las decisiones democráticas todavía pueden ser muy exactas – incluso si las estimaciones colectivas provienen de votantes mal informados, pero suficientemente diversos. Por lo tanto, una interpretación de estos hallazgos es que la polarización de opiniones, que puede derivar de la atracción hacia los extremos políticos (Goldenberg et al., 2023; Zimmerman et al., 2022), puede potencialmente mejorar el juicio democrático, como se propuso en trabajos previos (Shi et al., 2019).

Los estudios que examinan cómo la diversidad influye en el rendimiento grupal han proporcionado resultados disímiles, con efectos positivos (por ejemplo, Mohammed & Ringseis, 2001) y negativos (por ejemplo, Navajas et al., 2022) presentes dentro de la literatura de toma de decisiones colectivas (para una revisión, ver Sulik, Bahrami & Deroy, 2022). Esto sugiere que acumular evidencia diversa (Couch, 2022) puede no ser bueno o malo en sí mismo, sino que depende de la interacción con otras variables, como el tamaño del grupo (Pescetelli et al., 2021) y la estructura de la red (Baumann et al., 2024). Este trabajo aporta a esta literatura evidenciando que aumentar la diversidad puede reducir la exactitud a nivel individual, pero aumentarla a nivel colectivo. Crucialmente, se demuestra que este fenómeno ocurre en un dominio de pronóstico complejo, área en la que existe una larga tradición de estudiar el valor epistémico de combinar

predicciones diversas (Bates & Granger, 1969). Aunque este trabajo evidenció la existencia de este fenómeno, y estudió su robustez frente a diferentes métodos de agregación, trabajos futuros podrían optimizar aún más la sabiduría de multitudes extremizadas, por ejemplo, combinándola con estrategias de reducción de sesgo (debiasing) por medio de modelos generativos (Lee, & Danileiko, 2014; Lee et al., 2023).

## Capítulo 3

### **3. La Implementación Colectiva del Método de Fermi Subyace a la Sabiduría de Multitudes Deliberativas**

#### **3.1. Resumen**

Entender bajo qué condiciones grupos de personas presentan un mejor desempeño que individuos independientes es un problema clave en las ciencias sociales y psicológicas. Un concepto central en este campo es el de "sabiduría de las multitudes", según el cual la agregación de estimaciones independientes puede ser más exacta que el mejor juicio individual. Este fenómeno ha sido empleado en ámbitos variados, incluyendo política, salud, y negocios, y aplicado tanto en procesos de toma de decisiones como para realizar predicciones. Si bien en investigaciones previas se ha evidenciado que la deliberación grupal mejora la exactitud de las multitudes, los mecanismos detrás de esto aún no están claros. Este estudio aborda esa brecha, analizando la utilización grupal del "método de Fermi", que consiste en descomponer preguntas complejas en otras más simples, y combinar sus soluciones para mejorar las estimaciones. Un primer estudio con 520 participantes divididos en grupos de cuatro, interactuando en salas de chat, confirmó que las estimaciones grupales son más exactas que las individuales. Crucialmente, al analizar sus discusiones escritas (ya sea a través de evaluadores humanos o mediante un procedimiento automatizado), se confirmó además que los grupos que emplean el método de Fermi son aún más exactos. Un segundo estudio con 240 participantes proporcionó evidencia causal de esto, al pedirles explícitamente que utilizaran ya sea el método de Fermi, u otras estrategias alternativas. Un tercer estudio con 160 participantes respaldó la idea de que emplear

el método de Fermi colectivamente conduce a una mejora superior con respecto a utilizarlo individualmente. En suma, estos resultados ofrecen evidencia causal de que emplear colectivamente el método de Fermi mejora las estimaciones grupales, promoviendo su aplicación en estos dominios, ya sea mediante instrucciones explícitas, o mediante métodos de detección automatizados.

Como se mencionó en el Capítulo 1, el fenómeno de Sabiduría de Multitudes juega un rol central dentro de los problemas intelectivos. En el Capítulo 2 de esta tesis se analizó cómo optimizar este fenómeno en preguntas en las que no suele funcionar, pero preservando la independencia de los estimativos individuales (debido a que no se producía ningún tipo de interacción entre los participantes de cada experimento). La literatura previa ha encontrado resultados ambiguos al analizar el impacto de la influencia entre participantes, con algunos estudios evidenciando un efecto contraproducente de esta interacción (por reducción de independencia de los estimativos), y otros encontrando que por medio de la misma, se podían alcanzar respuestas más exactas. En este capítulo, se analizará el rol de las interacciones entre participantes en contextos de discusiones colectivas en grupos reducidos de participante.

### **3.2. Introducción**

La agregación de muchas estimaciones individuales a menudo supera a los estimativos provistos por expertos independientes (De Condorcet, 1785; Galton, 1907). Este fenómeno, popularmente conocido como la "sabiduría de las multitudes" (Surowiecki, 2005), se ha aplicado en una amplia gama de problemas, como la mejora de diagnósticos médicos (Kurvers et al., 2016), previsión de

eventos geopolíticos (Mellers et al., 2014), predicción de mercados financieros (Ray, 2006), ingeniería inversa del olor de las moléculas (Keller et al., 2017), y la verificación de hechos en las noticias (Allen, 2021), entre muchos otros. Comprender los mecanismos que subyacen a la sabiduría de las multitudes se ha convertido en un tema cada vez más relevante y crucial en las ciencias sociales y psicológicas (Kameda, Toyokawa y Tindale, 2022; Karachiwalla y Pinkow, 2021; Navajas et al., 2018; Kao y Couzin, 2014), dado su potencial para proporcionar información sobre cómo funcionan los grupos, y cómo optimizar sus rendimientos.

Muchas investigaciones previas han explorado el impacto de la influencia social en la sabiduría de las multitudes, arrojando resultados disímiles. Mientras que algunos estudios han informado efectos adversos de la influencia social (Raafat, Chater y Frith, 2007; Lorenz et al., 2011; Madirolas y de Polavieja, 2015), otros han demostrado que puede mejorar la exactitud colectiva (Gürçay, Mellers y Baron., 2015; Mellers, 2014; Bahrami, 2010; Juni y Eckstein, 2015). Por ejemplo, un estudio ha mostrado que la agregación de estimaciones colectivas de unos pocos grupos interactuantes puede superar ampliamente la sabiduría de muchos individuos (Navajas et al., 2018). Aunque este efecto ha sido replicado en entornos diversos, los procedimientos empleados por dichos grupos para llegar a estimaciones colectivas tan exactas aún restan por dilucidarse.

En Navajas et al. (2018) se analizó la posibilidad de explicar el bajo error de las estimaciones colectivas por medio de métodos de agregación de las estimaciones individuales iniciales. Sin embargo, a pesar de evaluar una amplia variedad de reglas de agregación, ninguna pudo explicar el bajo error mencionado. Además, cuando se les preguntó a los participantes por el procedimiento utilizado, los grupos manifestaron haber razonado y recalculado colectivamente

sus respuestas, para llegar a un consenso. Sin embargo, dos limitaciones importantes se desprenden de dicho análisis: 1) los grupos podrían haber combinado sus valores a través de una regla de agregación diferente de las estudiadas, y 2) al carecer de acceso a las transcripciones de esas conversaciones (orales), el trabajo anterior no pudo corroborar que los grupos hayan verdaderamente razonado de manera colectiva, recalculando la respuesta final. Asimismo, en el contexto de la estimación de variables, no está claro qué se entiende por "razonar colectivamente". Una posibilidad, ampliamente respaldada por la literatura previa, consiste en recalcular la variable de interés dividiendo el problema principal en problemas de estimación más pequeños, y realizando los cálculos necesarios para llegar a la respuesta final. Por ejemplo, para estimar el número de escalones en la escalera de un edificio grande, se podría suponer un número dado de pisos, una altura específica de un escalón individual, y una cierta cantidad de escalones en cada piso. Al combinar estas estimaciones independientes, es posible llegar a una estimación final del número total de escalones. Esta estrategia, comúnmente conocida como el "Método de Fermi", ha demostrado ser efectiva en una amplia variedad de ámbitos (Nityananda, 2014).

A la luz de los comentarios anteriores, en este trabajo se hipotetiza que el bajo error colectivo observado en Navajas et al. (2018) no se debe a una agregación de las respuestas individuales iniciales (H1), sino, en cambio, a que los participantes razonaron y recalcularon colectivamente la variable respectiva, mediante el Método de estimación de Fermi (H2). Para evaluar estas hipótesis, se realizó un estudio inicial para replicar los resultados en Navajas et al. (2018) en un contexto diferente, donde fue posible registrar las conversaciones y evaluar su contenido. Otros dos experimentos adicionales proveyeron evidencia convergente en favor de la utilización del

Método de Fermi en las discusiones grupales como una forma de reducir el error colectivo y, por lo tanto, mejorar la sabiduría de las multitudes.

### **3.3. Resultados**

#### **3.3.1. Experimento 1**

El experimento 1 involucró a 520 participantes (299 mujeres, edad promedio 25.3 años, desviación estándar 7.6 años), distribuidos en 130 grupos de cuatro individuos. Al igual que los demás estudios reportados en este trabajo, los participantes fueron reclutados en línea, y residían en Argentina en el momento del experimento. Los participantes tenían incentivos monetarios para estimar estas variables lo más exactamente posible (para más detalles, ver la sección Métodos).

El experimento constó de tres etapas (Fig. 1A). Durante la primera etapa (etapa i1), los participantes respondían individualmente a una serie de ocho preguntas de estimación de conocimientos generales (Tabla A6 en el Apéndice 2). En la segunda etapa (etapa c), fueron divididos en grupos de cuatro, y recibieron cuatro de las ocho preguntas anteriores (elegidas al azar). Se les instó a discutir estas preguntas en una sala de chat virtual, y se les pidió que intentaran llegar a un consenso. La última etapa (etapa i2) siguió el mismo procedimiento que la primera etapa. Todas las conversaciones escritas fueron registradas.

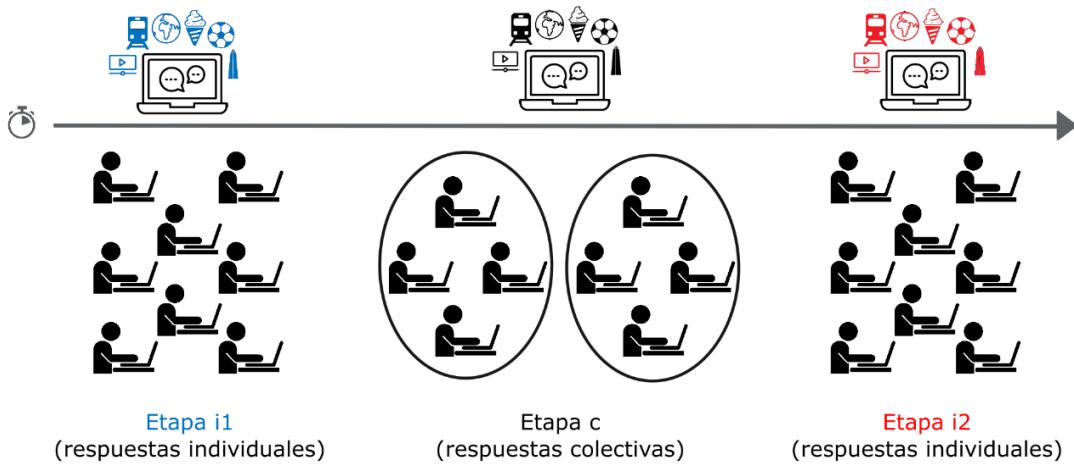
Los resultados presentados en la Figura 1 son consistentes con los hallazgos de Navajas et al. (2018), a pesar de las diferencias metodológicas (como el tamaño de la muestra, lugar del experimento, número de participantes por grupo, y el procedimiento de recolección de datos).

En primer lugar, al promediar las estimaciones proporcionadas por  $m$  grupos, siempre se obtiene un error colectivo más bajo que al promediar las estimaciones de  $n = 4.m$  individuos (Fig. 1B, ver Métodos para más detalles sobre el procedimiento de resampling utilizado). Esto es cierto independientemente del valor de  $m$ . Por ejemplo, el error colectivo de  $m = 8$  grupos fue sustancialmente más bajo que el error colectivo de  $n = 32$  individuos (test-t no apareado:  $t(498)=68.4$ ,  $p=2\times 10^{-257}$ ; tamaño del efecto: delta de Cohen = 6.12). A su vez, es posible comparar el promedio de las estimaciones individuales de cada grupo con la estimación colectiva respectiva de cada grupo (Fig. 1C). Al observar estas distribuciones de estimativos, se puede ver que el error colectivo de las estimaciones grupales es significativamente menor que el error colectivo del promedio de las estimaciones individuales de cada grupo (test-t apareado:  $t(124)=9.02$ ,  $p=3\times 10^{-15}$ ; tamaño del efecto: delta de Cohen = 0.96). Por último, si comparamos las estimaciones de ambas etapas individuales, separando según si las preguntas fueron discutidas en grupos o no, se vuelve evidente que hay una reducción del error significativamente mayor en el caso de las preguntas discutidas (Fig. 1D): mientras que los participantes mejoraron, en líneas generales, su exactitud en las preguntas no discutidas (test-t apareado:  $t(1667)=7.95$ ,  $p=3\times 10^{-15}$ ; tamaño del efecto: delta de Cohen = 0.09), la mejora para las preguntas discutidas fue ampliamente superior (test-t apareado:  $t(1161)=21.7$ ,  $p=3\times 10^{-88}$ ; tamaño del efecto: delta de Cohen = 0.66). Esto implica que discutir ciertas preguntas de manera grupal realmente llevó a los individuos a ser más exactos en esas preguntas.

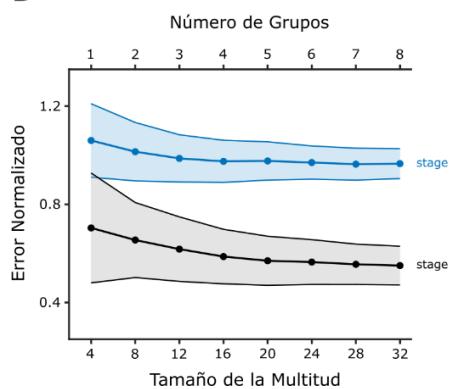
Una explicación plausible para estos resultados sería que esta mejora en la inteligencia colectiva (es decir, el hecho de que los grupos alcancen una respuesta más exacta de lo que se esperaría mediante la media aritmética) se debe a que las conversaciones son dominadas por

# Experimento 1 (N=520)

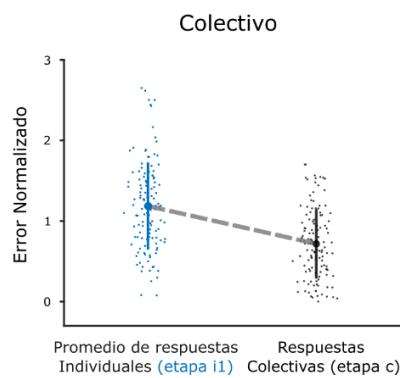
A



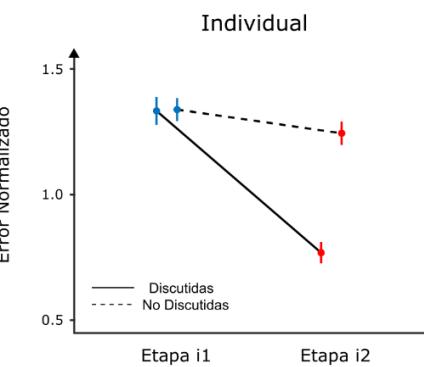
B



C



D



**Fig. 1. Resultados empíricos para el Experimento 1. (A)** El experimento tuvo tres etapas. En la Etapa i1, los participantes respondieron ocho preguntas de conocimiento general individualmente. En la Etapa c, los participantes se dividieron en grupos de cuatro y se les hizo la mitad de las preguntas de la etapa anterior. La Etapa i3 siguió el mismo procedimiento que la Etapa i1. **(B)** Error normalizado del promedio de  $n$  respuestas individuales (línea azul para la Etapa i1), y error normalizado del promedio de  $m = n/4$  estimaciones colectivas (línea negra, Etapa c). Las barras de error corresponden a la desviación estándar de las medias **(C)** Error normalizado promedio y distribuciones de errores normalizados de las respuestas individuales (azul) y las respuestas colectivas (negro). Las barras de error corresponden a la desviación estándar de las medias. **(D)** Error normalizado promedio para las etapas individuales, tanto para las preguntas discutidas como para las no discutidas. Las barras de error corresponden a los intervalos de confianza del 95%.

aquellos individuos que son más exactos o que tienen más confianza. Aunque nuestros datos muestran –razonablemente– que las respuestas grupales pueden explicarse como combinaciones lineales donde las estimaciones de personas con más exactitud y confianza tienen una influencia mayor, no hemos encontrado evidencia en los registros de chat de que esas personas hayan tenido más intervenciones (es decir, que hayan enviado más mensajes) o de que hayan usado un mayor número de palabras, en comparación con individuos con menos exactitud o confianza (ver Fig. A5 en el Apéndice 2 para más detalles).

En línea con las hipótesis presentadas en la Introducción, se contrataron diez evaluadores que fueron entrenados para distinguir entre la "estrategia de números" (compartir sus estimaciones individuales de la etapa anterior, y agregarlas) y la "estrategia de Fermi" (dividir el problema en partes más pequeñas, y recalcular su respuesta final). A continuación, los evaluadores proporcionaron calificaciones separadas sobre cuánto de cada estrategia estaba presente en cada conversación (Fig. 2A). Por ejemplo, en la conversación del lado izquierdo de la Figura 2A, los participantes se dedicaron principalmente a obtener y combinar sus respuestas individuales originales (comentarios resaltados en violeta), con sólo un comentario presente que hubiera implicado recalcular la respuesta mediante razonamiento colectivo (resaltado en azul). Como resultado, esta conversación debería tener una calificación alta para la estrategia de números y una calificación muy baja para la estrategia de Fermi. En cambio, en la conversación del lado derecho de la Figura 2A, los participantes apenas mencionaron sus estimaciones originales (comentario resaltado en violeta), pero dividieron el problema en tareas de estimación más pequeñas, y luego combinaron estas estimaciones para llegar a su respuesta final (comentarios

resaltados en azul). Por lo tanto, esta conversación debería tener una calificación muy baja para la estrategia de números y una calificación alta para la estrategia de Fermi.

A partir de las calificaciones proporcionadas por los evaluadores, se observa que, como era de esperar, hay una fuerte correlación negativa entre las calificaciones de Fermi y las calificaciones de Números (correlación de Pearson:  $r = -0.91$ ,  $p = 1 \times 10^{-170}$ , ver también Fig. 2B). Esto implica que los evaluadores tenían una buena comprensión de los conceptos involucrados, ya que sería poco probable encontrar muchas conversaciones donde se usaran ambas estrategias simultáneamente. Los evaluadores también fueron consistentes entre sí, ya que hubo una alta correlación entre sus respectivas calificaciones (Fig. 2C). Otro chequeo relevante (omitido de la Fig. 2 por brevedad) es que cuanto mayor era la calificación de Números para una conversación dada, mayor era el porcentaje de líneas en la conversación donde los participantes compartían sus respuestas de la etapa anterior. Lo contrario ocurre con las calificaciones de Fermi: cuanto más altas eran, menor era el porcentaje de líneas donde los participantes compartían sus respuestas individuales. Más importante aún, como se muestra en la Fig. 2D y 2E, hay una correlación positiva significativa entre el error normalizado y el promedio de las calificaciones de Números (correlación de Pearson:  $r = 0.16$ ,  $p = 0.004$ ); y hay una correlación negativa significativa entre el error normalizado y el promedio de las calificaciones de Fermi (correlación de Pearson:  $r = -0.21$ ,  $p = 2 \times 10^{-4}$ ). Esto implica que usar la estrategia de Fermi para abordar estas preguntas de sabiduría de las multitudes colleva a una mejora (es decir, una reducción en el error), mientras que utilizar la estrategia de Números resulta perjudicial (es decir, produce un mayor error, en comparación con la otra estrategia). Esto también se evidencia por la diferencia en el error colectivo entre los grupos que predominantemente emplearon la estrategia de Fermi (lo

que significa que su calificación de Fermi fue más alta que su calificación de Números) y los grupos que predominantemente emplearon la estrategia de Números (Fig. 2F). Si bien no hay una diferencia inicial en el error al promediar las respuestas individuales de cada grupo (test-t no apareado:  $t(324)=0.14$ ,  $p > 0.05$ ; tamaño del efecto: delta de Cohen = 0.017), sí existe una diferencia significativa entre las respuestas colectivas de los grupos que predominantemente emplearon la estrategia de Fermi, y aquellos que no lo hicieron (test-t no apareado:  $t(324)=3.20$ ,  $p = 0.0015$ ; tamaño del efecto: delta de Cohen = 0.38). También se debe destacar que, en ambos casos, la reducción en el error entre las respuestas colectivas de un grupo dado y el promedio de las respuestas individuales también es estadísticamente significativa ( $p = 0.045$  para los grupos con valor bajo en uso de la estrategia de Fermi, y  $p = 9 \times 10^{-16}$  para los grupos con valor alto), apoyando una vez más el uso de este método para mejorar la sabiduría de las multitudes clásica.

Comprender las mejores estrategias posibles para mejorar las discusiones grupales sobre preguntas fácticas –como las empleadas en este trabajo– es altamente relevante por varias razones. En primer lugar, los participantes podrían recibir instrucciones sobre qué estrategia usar para obtener mejores resultados. Como ejemplo, este método podría, potencialmente, ayudar a mejorar los mercados de decisiones en la obtención de mejores pronósticos. En el Experimento 2 se explora la posibilidad de instruir a los participantes sobre qué estrategia usar. En segundo lugar, dada la ubicuidad de las discusiones grupales en plataformas en línea –desde tablas de anuncios (“message boards”) y foros de discusión hasta comentarios en redes sociales–, disponer de un método para identificar qué estrategia se está utilizando (por ejemplo, una estrategia de “Números”) y sugerir a los participantes que la sustituyan por una estrategia mejor (como la estrategia de “Fermi”) sería una forma muy prometedora de mejorar la resolución de problemas

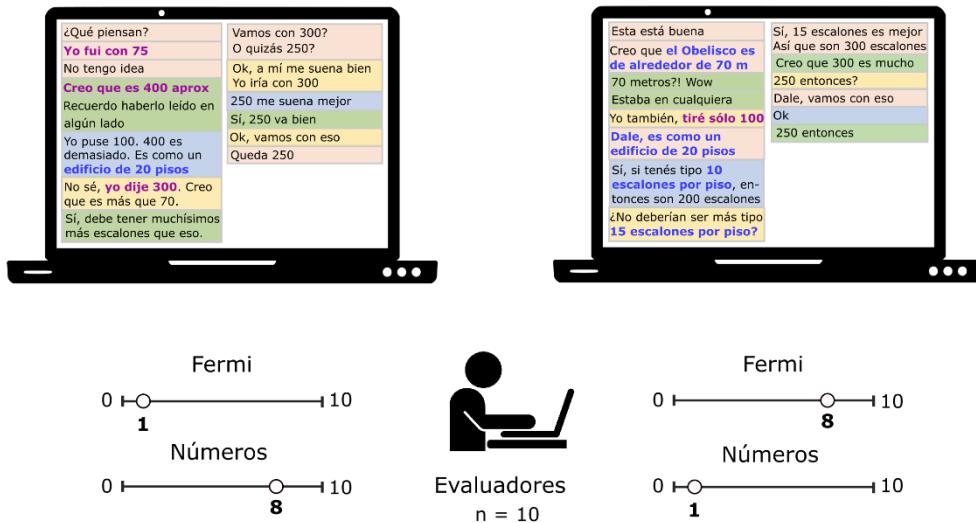
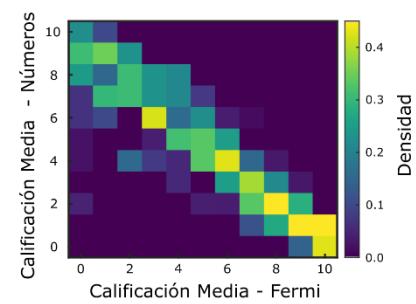
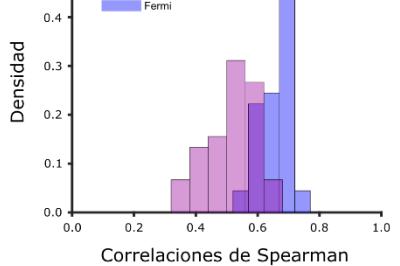
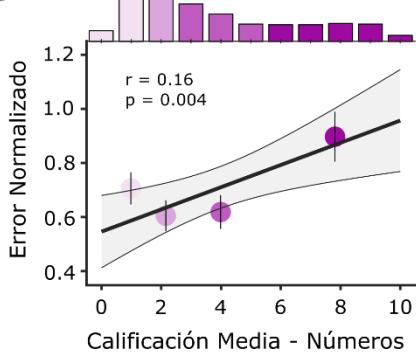
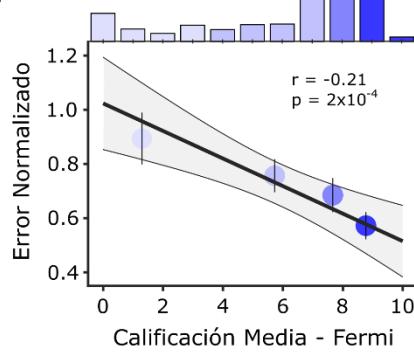
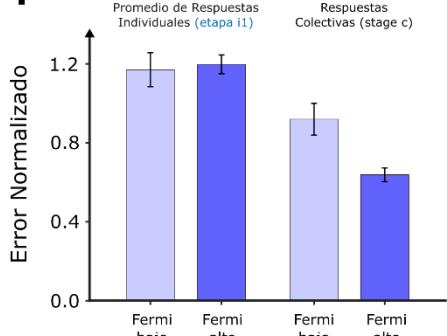
en línea. Sin embargo, hasta ahora, el método presentado aquí tiene una limitación importante: contratar evaluadores humanos es un método costoso, que consume mucho tiempo, no se produce en tiempo real, y no es escalable para categorizar conversaciones (sin mencionar la subjetividad inherente de las calificaciones humanas).

Con el fin de superar esta limitación, proponemos la siguiente intuición: más allá del método utilizado, si los grupos en verdad razonan colectivamente y recalculan la variable en cuestión, los participantes deberían usar palabras que sean semánticamente cercanas a las palabras de la pregunta correspondiente. Del mismo modo, si los grupos sólo están intercambiando sus respuestas anteriores, no debería ser necesario usar palabras que sean semánticamente cercanas a la pregunta realizada. Para cuantificar esta intuición, definimos un "puntaje de Fermi", obtenido aprovechando incrustaciones de palabras ("word embeddings") pre-entrenadas (Gutiérrez & Keith, 2019), específicamente a través de FastText, en idioma español (Bojanowski et al., 2017). Los word embeddings –representaciones vectoriales de palabras– poseen algunas propiedades útiles, siendo la más notable que palabras relacionadas semánticamente tienen vectores similares. Por ejemplo, la representación vectorial de "perro" debería ser relativamente similar a la de "gato", y menos similar cuando se compara con "transistor" (o cualquier otra palabra no relacionada). Los word embeddings son ampliamente utilizados en una gran variedad de dominios, como cuantificar el habla de los pacientes en psiquiatría computacional, o identificar signos de intoxicación, entre otras aplicaciones (Corcoran et al., 2018; Mikolov et al., 2017).

En este estudio se empleó FastText para medir la similitud semántica entre cada pregunta y cada conversación correspondiente. Específicamente, este método calcula la similitud entre

**A**

P: ¿Cuántos escalones tiene la escalera interna del Obelisco de Buenos Aires?

**B****C****D****E****F**

**Fig. 2. Análisis de Conversaciones.** (A) Ejemplo de conversaciones que conducen a calificaciones opuestas de las estrategias utilizadas en ellas. Diez evaluadores fueron entrenados en cómo clasificar conversaciones en términos de si los participantes agregaron sus respuestas individuales previas ("estrategia de Números") o alcanzaron una nueva respuesta colectiva utilizando el Método de Fermi ("estrategia de Fermi"). (B) Densidad de calificaciones para las estrategias "Fermi" y "Números". Las calificaciones para ambas estrategias estaban fuertemente correlacionadas en forma negativa: siempre que los evaluadores asignaban en promedio un valor alto para la estrategia de Fermi en una conversación, también tendían a asignar en promedio un valor muy bajo para la estrategia de Números en esa conversación. (C) Distribución de correlaciones entre los valores proporcionados por cada evaluador entre sí, para cada estrategia. La correlación de los valores proporcionados por diferentes evaluadores es generalmente muy alta, para ambas estrategias. (D) Error normalizado en función de las calificaciones de la estrategia de "Números" (promediadas entre evaluadores). Representamos los cuartiles de la

distribución de errores normalizados, junto con el mejor ajuste lineal de los datos, y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También mostramos el coeficiente de correlación de Pearson y el p-valor correspondiente. **(E)** Error normalizado en función de las calificaciones de la estrategia "Fermi" (promediadas entre evaluadores). Representamos los cuartiles de la distribución de errores normalizados, junto con el mejor ajuste lineal de los datos y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También mostramos el coeficiente de correlación de Pearson y el p-valor correspondiente. **(F)** Variación del error normalizado entre grupos y promedios de las respuestas individuales. Categorizamos un grupo para cada conversación como "bajo Fermi" cuando la calificación de Fermi era menor que la calificación de Números, o "alto Fermi" cuando la calificación de Fermi era mayor que la calificación de Números.

todas las palabras en la pregunta, y todas las palabras en la conversación, excluyendo palabras cortas (small words) y palabras vacías (stop words). Este proceso genera una distribución de valores de similitud (Fig. 3A). Dado que este es un proceso puramente automático, y podría utilizarse, en principio, en tiempo real, supera las limitaciones del método anterior de contratar y entrenar evaluadores humanos. Esta nueva calificación de similitud está correlacionada negativamente con el error normalizado (correlación de Pearson:  $r = -0.15$ ,  $p = 0.006$ ), lo que respalda aún más la idea de que discutir las variables relevantes (en lugar de sólo agrupar las estimaciones individuales originales) conduce a una reducción en el error colectivo (Fig. 3B). A su vez, la calificación de similitud también está correlacionada positivamente con la calificación de Fermi (correlación de Pearson:  $r = 0.22$ ,  $p = 5 \times 10^{-6}$ ), y correlacionada negativamente con la calificación de Números (correlación de Pearson:  $r = -0.16$ ,  $p = 0.001$ ), lo que implica que esta calificación captura de forma apropiada el empleo de la estrategia de Fermi.

Por último, se buscó evaluar si la calificación de similitud y su impacto en el error normalizado podían explicarse total o parcialmente por la calificación de Fermi, lo que indicaría que el método de Fermi está en efecto relacionado con esta medida automática. Para esto, se realizó un análisis de mediación (empleando 1000 muestras de bootstrap) con la calificación de Fermi actuando como mediador entre la calificación de similitud y el error normalizado (Fig. 3C). Descubrimos que el efecto directo de la primera en el último se reduce significativamente cuando se controla el efecto de la calificación de Fermi (efecto total =  $-0.11 \pm 0.03$ ,  $p = 0.006$ ; efecto directo =  $-0.08 \pm 0.04$ ,  $p = 0.02$ ; efecto indirecto =  $-0.03 \pm 0.01$ ,  $p = 0.009$ ). Esto respalda la idea de que la calificación de similitud está capturando el empleo de la estrategia de Fermi en las conversaciones.

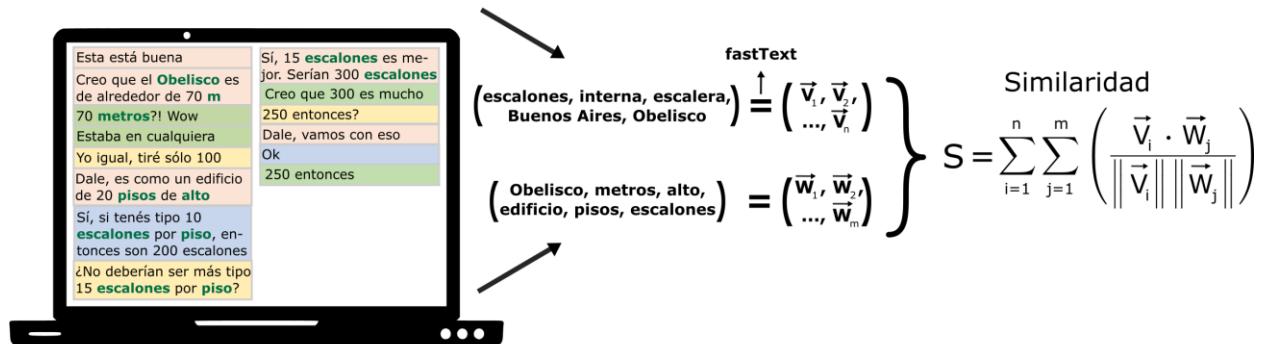
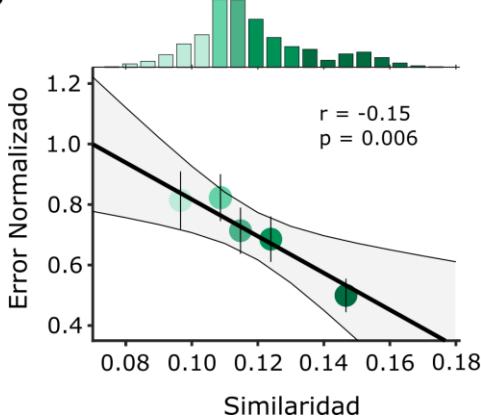
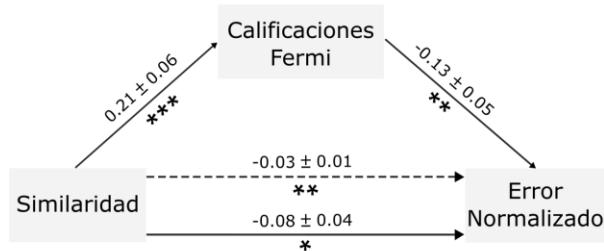
Hasta ahora, los resultados presentados aquí respaldan la existencia de una correlación entre el empleo del método de Fermi –cuantificado tanto con calificaciones proporcionadas por humanos como con un método automatizado– y la reducción del error colectivo en las discusiones grupales. Sin embargo, la naturaleza observacional de este estudio deja abierta la pregunta de si seguir de forma voluntaria la estrategia de Fermi influye causalmente en dicho error colectivo o no. Con el fin de evaluar esta relación causal, llevamos a cabo un segundo experimento, pre-registrado.

### **3.3.2. Experimento 2**

El experimento 2 fue pre-registrado en [https://aspredicted.org/TL2\\_Q1Q](https://aspredicted.org/TL2_Q1Q), e involucró a 240 participantes (148 mujeres, edad promedio 30.6 años, desviación estándar 9.8 años), distribuidos en 60 grupos de cuatro individuos. Una vez más, los participantes fueron reclutados en línea y

**A**

P: ¿Cuántos **escalones** tiene la **escalera interna** del **Obelisco de Buenos Aires**?

**B****C**

**Fig. 3. Análisis de Texto.** **(A)** Para cada conversación se calculó la similitud entre todas las palabras de la conversación con todas las palabras de la pregunta (excluyendo las palabras vacías y las palabras pequeñas). Con este fin, se generaron bolsas de palabras tanto para la pregunta como para la conversación, y se utilizó FastText para conseguir los vectores de palabras correspondientes. Luego, se calculó la similitud total entre todos los vectores de un grupo y todos los vectores del otro grupo. **(B)** Error normalizado en función de la similitud. Se representan los quintiles de la distribución de errores normalizados, junto con el mejor ajuste lineal de los datos y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También se exhibe el coeficiente de correlación de Pearson y el p-valor correspondiente. **(C)** Análisis de mediación entre similitud y error normalizado, con las calificaciones de Fermi (obtenidas de los evaluadores) actuando como mediador.

residían en Argentina en el momento del experimento. Los participantes tenían incentivos monetarios para estimar estas variables lo más exactamente posible (para más detalles, ver Métodos).

El experimento siguió el mismo procedimiento que el estudio anterior, con una excepción fundamental (Fig. 4A): antes de llegar a la etapa c (la etapa de deliberación colectiva), los participantes de cada grupo recibieron una instrucción en video, que los instaba a emplear o bien una estrategia de Números (agregando sus estimaciones de la etapa individual previa) o bien una estrategia de tipo Fermi (dividiendo el problema en problemas más pequeños, y luego recalculando la respuesta final). La asignación a cada condición fue aleatorizada (aunque todos los participantes en un grupo dado recibieron la misma instrucción). Todas las conversaciones escritas fueron registradas. Para más detalles y enlaces a las instrucciones en video, ver Métodos.

Se observó que el error normalizado se reduce significativamente para aquellos grupos que recibieron la instrucción de Fermi en comparación con los grupos que recibieron la instrucción de Números (Fig. 4B, test de suma de rangos de Wilcoxon,  $z=2.1$ ,  $p=0.04$ ; tamaño del efecto: delta de Cohen = 0.26).

Una vez más, se contrataron diez evaluadores para analizar el empleo de cada estrategia en todas las conversaciones de este experimento (siguiendo el mismo procedimiento que en el anterior). De manera consistente, la calificación de Números para los grupos que recibieron la instrucción de Números fue significativamente más alta que la misma calificación para los grupos que recibieron la instrucción de Fermi (Fig. 4C, prueba de suma de rangos de Wilcoxon,  $z=6.8$ ,  $p=1\times10^{-11}$ ; tamaño del efecto: delta de Cohen = 1.19). Lo opuesto resultó cierto para la calificación de Fermi (Fig. 4D, prueba de suma de rangos de Wilcoxon,  $z=6.3$ ,  $p=4\times10^{-10}$ ; tamaño del efecto:

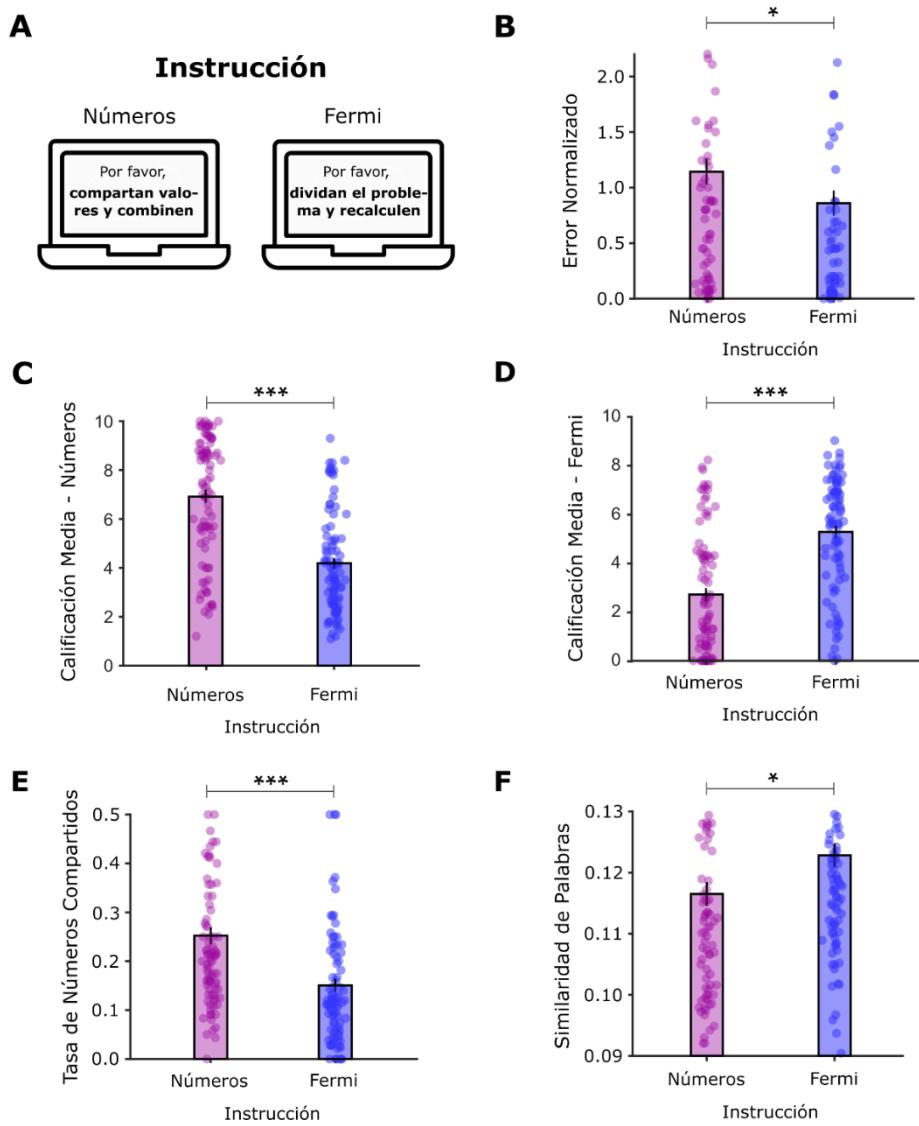
delta de Cohen = 1.06). Gráficos análogos a los de la Figura 3D y 3E pueden apreciarse en la Figura Suplementaria A6 en el Apéndice 2.

Otra forma de confirmar que los participantes siguieron las instrucciones adecuadamente es a través de la tasa de números compartidos de la etapa individual previa (etapa i1). En efecto, los grupos que recibieron la instrucción de Números compartieron más estimaciones de esa etapa en sus mensajes en comparación con los grupos que recibieron la instrucción de Fermi (Fig. 4E, prueba de suma de rangos de Wilcoxon,  $z=5.0$ ,  $p=6\times10^{-7}$ ; tamaño del efecto: delta de Cohen = 0.70).

Por último, de acuerdo con la calificación automática de similitud, los grupos que recibieron la instrucción de Números tuvieron una calificación de similitud más baja que los grupos que recibieron la instrucción de Fermi (Fig. 4F, prueba de suma de rangos de Wilcoxon,  $z=2.5$ ,  $p=0.01$ ; tamaño del efecto: delta de Cohen = 0.34). Esto es otra indicación de que los grupos realmente siguieron las instrucciones que recibieron, y que, en efecto, esto estuvo relacionado con la reducción en el error colectivo.

Si bien este experimento constituye otro paso adelante para comprender las variables que intervienen para mejorar la exactitud colectiva en discusiones de grupos pequeños, y si bien apoya la hipótesis de que los grupos emplean el método de Fermi para llegar a una respuesta más exacta, existe una explicación alternativa para este resultado: la reducción en el error debido al empleo del método de Fermi podría no estar relacionada con el hecho de que se lo esté utilizando *en grupos*. Este método también podría reducir el error cuando se lo utiliza individualmente, y no sólo en discusiones grupales. Para abordar esta cuestión, realizamos otro experimento pre-registrado.

## Experimento 2 (N=240)



**Fig. 4. Resultados empíricos del experimento 2.** (A) El procedimiento fue el mismo que el del experimento anterior, con una excepción: antes de la etapa c (deliberación colectiva), los participantes recibieron una instrucción de Números o una instrucción de Fermi, explicándoles e instándolos a emplear la estrategia correspondiente. (B) Error normalizado para los grupos que siguieron cada instrucción. (C) Calificación de Números proporcionada por evaluadores humanos para los grupos que siguieron cada instrucción. (D) Calificación de Fermi proporcionada por evaluadores humanos para los grupos que siguieron cada instrucción. (E) Tasa de números compartidos de la primera etapa individual que estuvieron presentes en las discusiones, para los grupos que siguieron cada instrucción. (F) Calificación de similitud para los grupos que siguieron cada instrucción.

### **3.3.3. Experimento 3**

El experimento 3 fue pre-registrado en [https://aspredicted.org/SZ5\\_MNP](https://aspredicted.org/SZ5_MNP), e involucró a 160 participantes (95 mujeres, edad promedio 29.8 años, desviación estándar 10.9 años), 80 de los cuales fueron divididos en 20 grupos de 4 individuos, y los 80 restantes realizaron todo el procedimiento de forma individual. Una vez más, los participantes fueron reclutados en línea y residían en Argentina en el momento del experimento. Los participantes tenían incentivos monetarios para estimar estas variables lo más exactamente posible (para más detalles, ver Métodos).

El experimento siguió el mismo procedimiento que el estudio anterior, con dos diferencias fundamentales:

- 1) La mitad de los participantes fueron asignados aleatoriamente a una condición de deliberación colectiva, que siguió exactamente el mismo procedimiento que la condición de instrucción de Fermi del Experimento 2 (es decir, se les instó a seguir el método de Fermi y deliberar en grupos, ver Fig. 5A).
- 2) La otra mitad no fue dividida en grupos, sino que recibió una instrucción de Fermi modificada que los instruía a aplicar el método de Fermi individualmente, y escribir su procedimiento (Fig. 5B).

Todas las conversaciones escritas fueron registradas. Para más detalles y enlaces a las instrucciones en video, ver Métodos.

Este procedimiento permitió comparar eficazmente el empleo del método de Fermi como estrategia para mejorar la exactitud, tanto en forma individual como en grupos. Con este fin, se

comparó la reducción en el error entre la primera etapa individual (etapa i1) y la última etapa individual (etapa i2), tanto para preguntas discutidas como no discutidas. A partir de los Experimentos 1 y 2 (por ejemplo, Fig. 1D), ya resulta evidente que este procedimiento es efectivo para mejorar la exactitud individual en la etapa i2. Este efecto fue replicado (Fig. 5C), encontrando, una vez más, una reducción significativamente mayor en el error para el caso de preguntas discutidas. Mientras que los participantes en general mejoraron su exactitud en las preguntas no discutidas (test-t pareado:  $t(275)=2.76$ ,  $p=0.006$ ; tamaño del efecto: delta de Cohen = 0.08), la mejora para las preguntas discutidas fue vastamente superior (test-t pareado:  $t(194)=7.78$ ,  $p=4\times 10^{-13}$ ; tamaño del efecto: delta de Cohen = 0.60). Críticamente, al observar la condición de deliberación individual (Fig. 5D), también vemos una reducción muy similar en el error para las preguntas no discutidas (test-t pareado:  $t(219)=2.57$ ,  $p=0.009$ ; tamaño del efecto: delta de Cohen = 0.06), que es menor que la reducción en el error para las preguntas discutidas (test-t pareado:  $t(219)=5.96$ ,  $p=1\times 10^{-8}$ ; tamaño del efecto: delta de Cohen = 0.27). Sin embargo, esta última reducción en el error en la condición de deliberación individual es mucho menor que la reducción en la condición de deliberación grupal (test-t no pareado:  $t(413)=5.92$ ,  $p=7\times 10^{-9}$ ; tamaño del efecto: delta de Cohen = 0.58). Esto se puede ver fácilmente al comparar los paneles C y D de la Figura 5.

Por último, calculamos la calificación de similitud para ambas condiciones, y analizamos el cambio en el error normalizado con respecto a esta calificación. En la condición de deliberación colectiva (Fig. 5E), encontramos que estas variables están correlacionadas negativamente (correlación de Pearson:  $r = -0.19$ ,  $p = 0.007$ ). Lo mismo ocurre para la condición de deliberación individual (Fig. 5F, correlación de Pearson:  $r = -0.20$ ,  $p = 0.004$ ). Esto implica que, una vez más, la

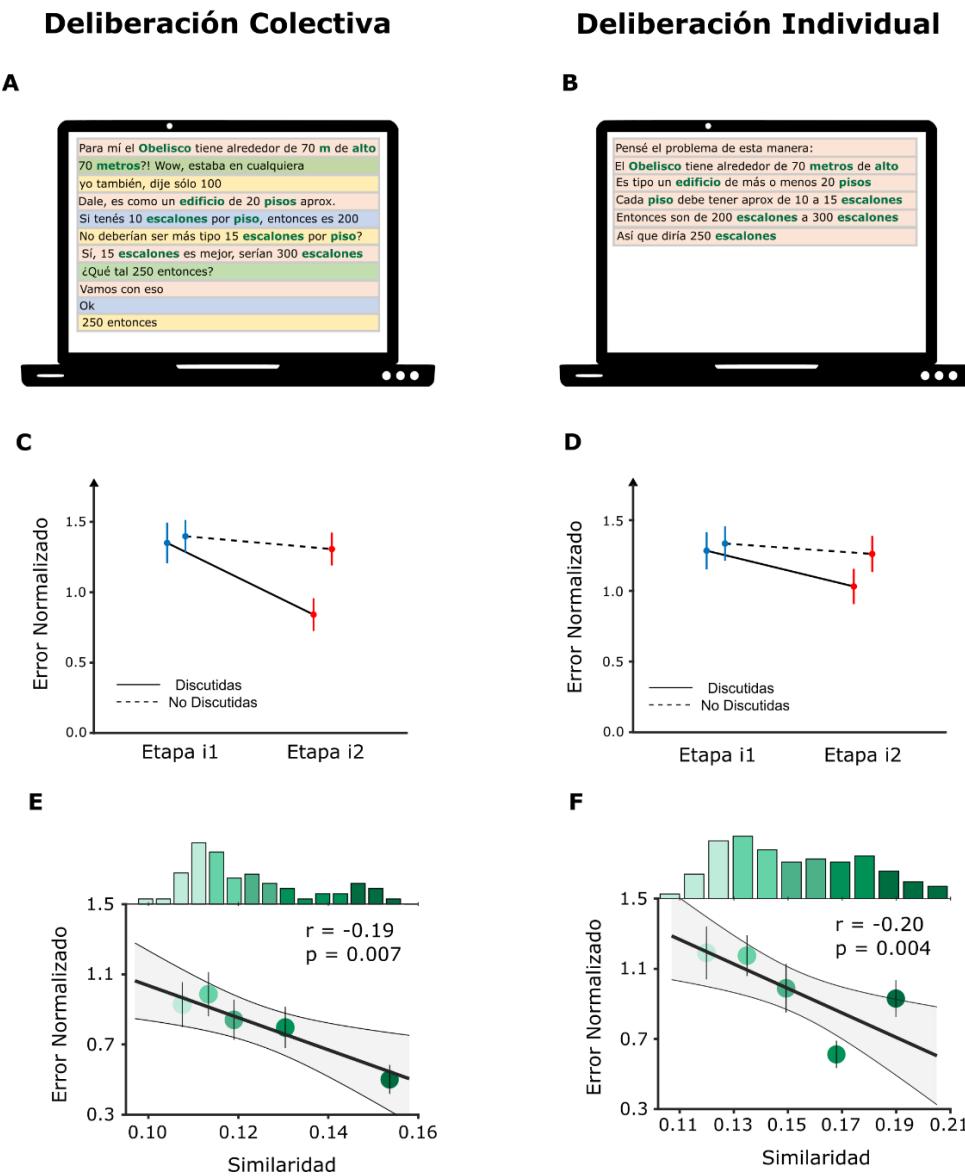
calificación de similitud logra capturar el empleo del método de Fermi, y podría ser potencialmente útil como un medio para detectar en forma automática grupos e individuos que no lo están utilizando (y que podrían beneficiarse de ser instados a hacerlo, derivando en una reducción en el error colectivo).

### **3.4. Discusión General**

En este trabajo se presenta evidencia que respalda el uso de una estrategia de razonamiento colectivo basada en el conocido “Método de Fermi” como herramienta para aumentar la exactitud colectiva en discusiones de grupos pequeños. Partiendo de literatura previa que muestra que la agregación de estimaciones colectivas de unos pocos grupos interactuantes supera ampliamente la sabiduría de las multitudes tradicional (Navajas, 2018), se realizaron tres experimentos conductuales que proporcionan evidencia convergente a favor de esta nueva estrategia, permitiendo mejorar aún más el procedimiento descrito en dicho trabajo. También se presenta una estrategia automatizada para detectar el uso del Método de Fermi en conversaciones escritas, con implicancias tanto para la clasificación de conversaciones previamente mantenidas como para el desarrollo potencial de intervenciones en tiempo real para mejorar las conversaciones en curso.

Se evidenció que los grupos que utilizan el Método de Fermi son más exactos que los grupos que simplemente agregan sus estimaciones previas (Experimento 1), y que incentivar a los grupos a usar dicho método reduce su error colectivo (Experimento 2). También se corroboró que este efecto es mucho más fuerte cuando la estrategia se emplea en grupos, en comparación con el

## Experimento 3 (N=160)



**Fig. 5. Resultados empíricos del experimento 3.** El procedimiento fue el mismo que en el experimento anterior, con dos excepciones: **(A)** la mitad de los participantes fueron divididos en grupos y recibieron la instrucción original de Fermi, aplicando el método de Fermi de manera grupal. **(B)** La mitad restante procedió individualmente y recibió una instrucción modificada que les indicaba que también aplicaran el método de Fermi, pero de manera individual (escribiendo su procedimiento). **(C)** Error normalizado promedio para las etapas individuales, tanto para preguntas discutidas como no discutidas, para los participantes que deliberaron colectivamente. Las barras de error corresponden a los intervalos de confianza del 95%. **(D)** Error normalizado promedio para las etapas individuales para los participantes que

deliberaron individualmente. Se siguió el mismo procedimiento que para el panel anterior. **(E)** Error normalizado en función de la similitud para los participantes que deliberaron colectivamente. Se represetaron los quintiles de la distribución de errores normalizados, junto con el mejor ajuste lineal de los datos y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También se exhibe el coeficiente de correlación de Pearson y el p-valor correspondiente. **(F)** Error normalizado en función de la similitud para los participantes que deliberaron individualmente. Se siguió el mismo procedimiento que para el panel anterior.

uso individual del Método de Fermi para estimar estas variables (Experimento 3). Adicionalmente, se desarrolló un algoritmo de clasificación automatizado –basado en calcular la similitud semántica entre las conversaciones escritas y la pregunta realizada– que evaluaba el uso de este método, obteniendo resultados en concordancia con la clasificación manual de evaluadores humanos capacitados. En líneas generales, estos hallazgos implican que es posible aumentar la exactitud de grupos que resuelven preguntas de estimación complejas al incentivarlos a emplear el Método de Fermi, y que sería viable hacerlo de manera automatizada.

Si bien hasta ahora nuestro procedimiento ha sido probado sólo en conversaciones escritas, la replicación de los resultados de Navajas et al. (2018) en este contexto particular implica que es muy probable que el Método Fermi también se utilice en discusiones orales. Además, el método automatizado desarrollado en este trabajo podría extenderse a comunicaciones orales de manera relativamente directa, una línea de investigación muy prometedora que dejamos abierta para trabajos futuros.

El primer estudio sobre la sabiduría de las multitudes fue considerado como una demostración empírica de que las reglas de agregación democráticas pueden ser confiables y eficientes (Galton, 1907). Desde entonces, la mayoría de los intentos por aumentar la sabiduría colectiva se han

basado en la idea de que algunas opiniones son más meritorias que otras, y se ha propuesto encontrar aquellas opiniones más exactas mediante algún algoritmo de ponderación no uniforme (Madirolas & de Polavieja, 2015; Mannes, Soll, & Larrick, 2015; Prelec, Seung, & McCoy, 2017). Una notable excepción fue el trabajo de Navajas et al. (2018), donde los autores mostraron que la sabiduría de las multitudes podría aumentarse mediante simples discusiones cara a cara dentro de grupos, complementadas con un muestreo entre grupos. Su idea principal era que la recopilación de conocimiento de individuos que participaron en debates independientes reduce el error colectivo (sin comprometer el principio democrático de "un voto, un valor"). Sin embargo, al no tener acceso a un registro de las conversaciones reales, solo podían hacer suposiciones sobre cómo estas discusiones grupales mejoraban la sabiduría de las multitudes. Este trabajo contribuyó a llenar este vacío mostrando que la mejor estrategia no consiste en utilizar algoritmos de agregación –como se planteaba en la mayor parte de la literatura previa–, sino que los grupos más exactos emplean una estrategia conocida como el Método de Fermi, en la que dividen el problema central en problemas más pequeños, los resuelven, y luego realizan los cálculos necesarios para llegar a su respuesta final. Nuestros resultados parten de la noción política básica de que las encuestas deliberativas son un mecanismo práctico para resolver el dilema entre la igualdad y la deliberación. Navajas et al. (2018) demostraron que, en cuestiones de conocimiento general, la agregación de consensos obtenidos en grupos pequeños aumenta la sabiduría de las multitudes. Aquí se mostró que es posible ir un paso más allá, al instruir a los grupos a emplear el Método de Fermi, mejorando así su exactitud colectiva. Además, se evidenció que es posible utilizar un algoritmo automatizado para evaluar el uso de este método, lo que potencialmente permitiría el desarrollo de nuevas intervenciones que podrían ayudar a

dirigir a los grupos deliberativos hacia mejores estimaciones colectivas. Estos resultados respaldan las teorías políticas que postulan que la deliberación auténtica, y no simplemente la votación, puede llevar a mejores decisiones democráticas (Cohen, 1997).

### **3.5. Métodos**

#### **3.5.1. Desarrollo de una aplicación web para experimentos de decisiones grupales**

Esta sección proporciona una descripción detallada de la plataforma innovadora que fue diseñada y utilizada para los tres estudios presentados. Dicha plataforma digital permitió a los participantes intervenir activamente en debates, y responder preguntas utilizando una interfaz de sala de chat de uso sencillo. Para construir la plataforma, utilizamos un marco web basado en Python llamado Django Framework. Nuestro proyecto también utilizó otras librerías, incluyendo Django Channels (que extiende las capacidades de Django más allá de HTTP y permite trabajar con WebSockets, una tecnología ideal para implementar salas de chat). Todos los códigos necesarios y detalles adicionales requeridos para poder utilizar, construir o modificar esta plataforma están disponibles públicamente en <https://github.com/facuzeta/chat-room-platform>. Lanzamos la plataforma utilizando el servicio de orquestación Elastic Beanstalk, proporcionado por Amazon Web Services.

Para llevar a cabo los estudios mencionados anteriormente, se utilizó el siguiente protocolo:

- Antes de comenzar el experimento correspondiente, todos los participantes se unieron a una videollamada (Google Meet) donde se les comunicaron instrucciones detalladas para

el uso de la plataforma, y la realización del experimento. Luego, se les envió un correo electrónico con un enlace para acceder a la plataforma, que contenía todas las instrucciones necesarias, así como el consentimiento informado.

- Una vez que los participantes ingresaron a la plataforma, una pantalla de bienvenida los recibió, donde completaron su nombre, edad, género y nivel educativo más alto.
- Luego, se dio a los participantes la tarea de leer y responder una serie de preguntas. En todos los estudios, estas involucraban la estimación de cantidades inciertas. También se requería que los participantes expresaran el grado de confianza en sus respuestas. Disponían de un tiempo limitado para responder, y un temporizador les mostraba cuánto tiempo les quedaba.
- A continuación, se les pidió a los individuos que discutieran algunas de estas preguntas en grupo a través de texto escrito en un chat, y se les instó a *intentar llegar a un consenso*. Si esto sucedía, cada uno de ellos debía ingresar la respuesta colectiva que habían alcanzado. Un temporizador les mostraba cuánto tiempo tenían para responder a cada pregunta específica.
- Por último, los participantes volvieron a leer y responder la misma serie de preguntas original. Una vez más, disponían de un tiempo limitado para responder a todas ellas, y un temporizador les mostraba cuánto tiempo les quedaba.

### **3.5.2. Participantes y Preguntas**

Los participantes fueron informados de que su participación era completamente voluntaria, y que podían retirar su participación en cualquier momento. Todos los datos fueron

completamente anónimos. El protocolo experimental fue aprobado por el comité de ética de la Universidad Abierta Interamericana (Buenos Aires, Argentina), protocolo 0-1108.

El experimento 1 involucró a 520 participantes (299 mujeres, edad media 25.3 años, desviación estándar 7.6 años), distribuidos en 130 grupos de cuatro individuos. El experimento 2 fue pre-registrado ([https://aspredicted.org/TL2\\_Q1Q](https://aspredicted.org/TL2_Q1Q)) y se llevó a cabo con datos recopilados de 60 grupos de 4 individuos, totalizando 240 participantes (148 mujeres, edad media 30.6 años, desviación estándar 9.8 años). El experimento 3 también fue pre-registrado ([https://aspredicted.org/SZ5\\_MNP](https://aspredicted.org/SZ5_MNP)) e involucró a 160 participantes (95 mujeres, edad media 29.8 años, desviación estándar 10.9 años), 80 de los cuales se dividieron en 20 grupos de 4 individuos, y los 80 restantes realizaron todo el procedimiento individualmente. En los tres estudios, los participantes fueron reclutados en línea y residían en Argentina en el momento del experimento.

En todos los casos, los participantes fueron incluidos en un sorteo de cinco premios de \$AR 20,000 cada uno (aproximadamente 100 USD). Este procedimiento llevó a una tarifa de participación esperada de aproximadamente 4 USD por hora de experimento, lo cual es una tarifa justa basada en las pautas de participación de sujetos en la Universidad Torcuato Di Tella en Buenos Aires, Argentina.

En el experimento 1 se instó a los participantes a "intentar llegar a un consenso". En el experimento 2, además de esto, los participantes fueron asignados a una de dos condiciones. En la condición de "Números", se les instruyó explícitamente a compartir sus estimaciones numéricas iniciales. En la condición "Fermi", se les instruyó explícitamente a dividir el problema en problemas de estimación más pequeños, y realizar todos los cálculos relevantes para llegar a

su respuesta final. En ambos casos, estas instrucciones se proporcionaron a través de los siguientes videos:

- Video de instrucciones de la condición Números:

<https://www.youtube.com/watch?v=8VAuRjo882M>

- Video de instrucciones de la condición Fermi:

<https://www.youtube.com/watch?v=KHgEh2qg0Mk>

En el Estudio 3, todos los participantes recibieron las mismas instrucciones que en la condición Fermi del estudio anterior, pero la mitad de ellos fueron divididos en grupos de cuatro, y la mitad restante procedió individualmente. Al primer grupo se le pidió que "intentara llegar a un consenso". Al segundo grupo se le pidió que utilizaran el Método de Fermi por sí mismos, escribiendo su procedimiento en la misma sala de chat que se utilizó para el primer grupo (pero permaneciendo solos durante todo el experimento).

En los tres experimentos, las preguntas fueron presentadas en orden aleatorio en las primeras dos etapas, mientras que en la etapa 3 se repitió el orden que habían presentado en la etapa 1.

La selección de preguntas se basó en los siguientes criterios:

- La respuesta podría obtenerse de referencias confiables y de confianza.
- Las preguntas permitían un tiempo de estimación razonable de 5 minutos.
- Los valores asociados con las preguntas no eran excesivamente grandes en escala.

Para una lista completa de las preguntas, ver tabla S1 en Información Suplementaria.

### **3.5.3. Procedimiento**

En la primera etapa del experimento, los participantes respondieron de manera independiente a una serie de ocho preguntas de estimación numérica dentro de un límite de tiempo de 5 minutos. Junto con sus respuestas, indicaron su nivel de confianza, utilizando una barra deslizante que iba desde “bajo” hasta “alto”. En la segunda etapa los participantes se comunicaron a través del chat para intentar llegar a un consenso sobre los valores de cuatro de las ocho preguntas de la primera etapa. Cada problema discutido en esta etapa tenía un límite de tiempo de 5 minutos. Si se lograba un consenso antes, los participantes tenían que esperar hasta que pasaran los 5 minutos para poder pasar al siguiente problema. Durante esta etapa, no se les pidió a los participantes que informaran sus niveles de confianza. La tercera etapa del experimento fue similar a la primera etapa: los participantes abordaron individualmente las mismas preguntas de estimación numérica, dentro de un límite de tiempo de cinco minutos. Si todos los participantes completaban esta etapa antes de tiempo, entonces se los dirigía al mensaje de despedida (“Gracias”).

El experimento 1 siguió el procedimiento previamente mencionado, sin modificaciones adicionales. El experimento 2 también siguió dicho procedimiento, con la adición de una instrucción sobre qué estrategia debería usar el grupo (ya sea la estrategia “Fermi” o la estrategia “Números”) antes del inicio de la segunda etapa. El experimento 3 siguió el procedimiento del experimento 2, con dos modificaciones:

- Todos los participantes recibieron la instrucción para emplear la estrategia Fermi.
- La mitad de los participantes fue dividida en grupos, y estos procedieron exactamente como en el experimento 2; la otra mitad recibió instrucciones para emplear la estrategia “Fermi”

individualmente, y escribir su procedimiento. Estos últimos no fueron divididos en grupos, sino que completaron la segunda etapa de forma individual.

### **3.5.4. Exclusiones**

En la pregunta 3, algunos participantes respondieron en una escala diferente a la solicitada; para resolver este problema, se eliminaron todas las respuestas que fueran mayores a mil millones. También se eliminaron las respuestas con un error absoluto que fuera mayor o menor a 2.5 desviaciones absolutas medias (MAD) de la mediana (en cualquiera de las tres etapas de cada experimento).

## Capítulo 4

### 4. Multitudes Polarizadas: Consenso y Bipolarización en un Modelo de Argumentos Persuasivos

#### 4.1. Resumen

Comprender la dinámica de formación de opinión en sistemas sociales es de gran relevancia en diversos contextos sociales. En particular, es relevante para la deliberación política, y otros procesos de toma de decisiones grupales. Si bien investigaciones previas han presentado diferentes enfoques para modelar las dinámicas sociales, la mayoría de estos se centran en mecanismos de interacción donde los individuos modifican sus opiniones en función de las opiniones de los otros, sin invocar un mecanismo latente de argumentación. En este capítulo se presenta un modelo donde los cambios de opinión se deben a intercambios explícitos de argumentos, y se analizan los estados colectivos emergentes en términos de reglas dinámicas simples. Se observa que, cuando las interacciones son equiprobables y simétricas, el modelo sólo presenta soluciones de consenso. Sin embargo, cuando se incluyen la homofilia, el sesgo de confirmación, o ambos, surge y predomina la bipolarización. Esto se fundamenta en que los individuos sesgados ya no pueden aceptar información contraria de sus oponentes durante los intercambios de argumentos. En todos los casos, la preponderancia de cada estado estable depende de la relación entre el número de agentes y el número de argumentos disponibles en la discusión. En líneas generales, este capítulo describe la dinámica y evidencia las condiciones en las que se espera que los agentes deliberativos construyan sociedades polarizadas.

Los Capítulos 2 y 3 de esta tesis estuvieron relacionados con problemas intelectivos (como se describió en el Capítulo 1). El Capítulo 4 aborda la temática de problemas no intelectivos (de opinión). Se trata de un análisis teórico basado en modelado computacional basado en agentes, donde se describen los estados macroscópicos de un sistema formado por muchos agentes que interactúan intercambiando argumentos, y modificando sus opiniones de manera acorde. De esta manera, es posible profundizar en una rama de problemas que no necesariamente presentan respuestas correctas, pero que resultan de gran interés para las sociedades democráticas actuales.

## **4.2. Introducción**

La mayoría de las personas participa diariamente de diferentes tipos de debates y discusiones. Con frecuencia, en este tipo de interacciones, se invocan e intercambian argumentos, lo que eventualmente puede llevar a que se modifiquen las opiniones iniciales, así como las decisiones a tomar. Debates políticos (como los debates presidenciales), referendos (como el del "Brexit" de 2016), grupos de discusión (paneles médicos, jurados, etc.) e incluso reuniones laborales, son ejemplos de discusiones –en escalas muy diferentes–, en las que se deben tomar decisiones importantes. Estas incidencias guiaron el desarrollo de numerosos estudios experimentales y teóricos, centrados en comprender la dinámica de formación de opinión. Desde un punto de vista teórico, la física estadística proporciona herramientas cuantitativas muy poderosas, particularmente útiles para el estudio de sistemas complejos compuestos por muchos agentes interactuantes –como se puede apreciar en varias revisiones recientes y exhaustivas (Baronchelli, 2018; Castellano et al., 2009; Helbing, 2010; Weidlich, 1991).

En este marco de referencia, los **modelos basados en agentes** constituyen una técnica ampliamente difundida para derivar estados macroscópicos a partir de interacciones microscópicas simples entre agentes interactuantes. Se han aplicado eficientemente para estudiar diversos fenómenos y entornos, como el comportamiento y el tráfico de peatones (Castellano et al., 2009), la agricultura (Berger, 2001), la economía (Farmer & D. Foley, 2009), los sistemas socio-ecológicos (Filatova et al., 2013), la demografía (Billari & Prskawetz, 2012), entre muchos otros (Castellano et al., 2009; Goldstone & Janssen, 2005; Kohler & Gumerman, 2000).

Generalmente, los esfuerzos pasados en la confección de modelos de dinámica de formación de opinión basados en agentes involucraron el uso de opiniones discretas –como los modelos de votantes (Clifford & Sudbury, 1973; Holley & Liggett, 1975; Liggett, 2012; Cox & Griffeath, 1976; Krapivsky, 1992; Frachebourg & Krapivsky, 1996), modelos de Sznajd (Sznajd-Weron, 2005; Chmiel et al., 2017; Schneider, 2004; Martins, 2008; Crokidakis et al., 2014; Schwammle et al., 2007; Travieso & da Fontoura Costa, 2006) o modelos de reglas de mayoría, es decir, “majority rules” (Galam, 1986; Galam, 2002; Tessone et al., 2004; Schelling, 2006; Granovetter & Soong, 1986). En cambio, el uso de opiniones continuas dio lugar a una serie de modelos diferente, comenzando con el modelo original de DeGroot (DeGroot, 1974). Allí, la coexistencia de opiniones macroscópicas diferentes se introdujo mediante el concepto de **confianza acotada** (Deffuant et al., 2000). Este implicó que los agentes interactuantes ignoraban las opiniones de los demás agentes cuando estas eran extremadamente diferentes. Este tipo de modelos fueron estudiados extensivamente en los años recientes (Hegselmann & Krause, 2002; Kurahashi-Nakamura et al., 2016; Lorenz, 2007; Hegselmann & Krause, 2005; Urbig & Lorenz, 2007; Lorenz et al., 2020).

Existen también algunas contribuciones que combinan ambos enfoques –opiniones continuas y discretas. Por ejemplo, un modelo para opiniones interpretadas como acciones discretas, pero representadas internamente como una opinión continua, se presentó en Martins (2008). Se adoptó un enfoque similar en Balenzuela et al. (2015), donde las opiniones externas discretas emergen de posturas internas subyacentes. La dinámica de este modelo imita un proceso de recopilación de información: la postura de un agente puede cambiar después de la interacción con otro agente y, a su vez, esto puede inducir un cambio de opinión, si su postura cruza un umbral específico. Así, la opinión externa representa, de alguna manera, una decisión binaria específica, y la postura interna refleja cuán convencido está un agente acerca de la opinión manifestada. Aunque se desarrollaron previamente otros modelos basados en la acumulación de información, no involucran la combinación entre opiniones externas discretas y posturas internas continuas mediadas por umbrales.

La dinámica de opiniones produce una serie de estados macroscópicos que ocurren naturalmente en nuestra sociedad (Lorenz et al., 2020). Dada su relevancia política, económica y social, estos estados han sido estudiados ampliamente (Baronchelli, 2018; Burnstein and A. Vinokur, 1977; Sunstein, 2002; Jager & Amblard, 2005; Sampedro & Pérez, 2008; Dandekar et al., 2013; Jönsson et al., 2015; Flache et al., 2017). El **consenso** (cuando todos los miembros del grupo adhieren a la misma opinión) y la **bipolarización** (cuando se forman dos grupos distintos y polarizados) son ejemplos relevantes que se encuentran comúnmente en temas con afirmaciones binarias (“a favor” o “en contra”, por ejemplo). Aunque la bipolarización puede ser inducida introduciendo un mecanismo de influencia negativa (distanciamiento intencional de agentes disímiles), la evidencia que respalda su existencia es insuficiente y controvertida (Krizan

& Baron, 2007). Esto ha llevado a la propuesta de otros mecanismos capaces de reproducir dicho estado colectivo. Por ejemplo, en Mas & Flache, (2013), los autores proponen un modelo para alcanzar bipolarización en ausencia de influencia negativa, mediante la introducción de un intercambio explícito de argumentos, junto con la homofilia (mayor probabilidad de interacción con agentes similares en lugar de disímiles). Sus suposiciones se basan en la Teoría de Argumentos Persuasivos (PAT), que establece que los cambios de opinión se derivan de intercambios de argumentos, y dependen tanto del número, como de la fuerza (o persuasión) de los mismos.

En este capítulo se presenta un modelo basado en agentes donde los estados de los agentes están representados por una *opinión* discreta externa, con *posturas* internas continuas subyacentes. Sin embargo, las interacciones se dan mediante un intercambio explícito de argumentos, basado en PAT. Se desarrollan un conjunto de reglas para este intercambio, y se explora el nuevo espacio de parámetros, introduciendo también modificaciones destinadas a modelar ciertos sesgos cognitivos reportados previamente en la literatura, que han sido vinculados a los estados macroscópicos observados en la sociedad. Este capítulo está organizado de la siguiente manera: en la sección *Modelo* se describe la implementación del modelo, y el fundamento teórico en el que se basa. En la sección *Resultados* se presentan los resultados principales con respecto a los estados colectivos producidos por el modelo en la versión imparcial (subsección *Modelo Imparcial*) y en la versión sesgada, con sesgo de confirmación (subsección *Modelo con Sesgo de Confirmación*), y homofilia (subsección *Modelo con Homofilia*). Finalmente, en la sección *Conclusiones* quedan delineadas las principales conclusiones y se discuten posibles modificaciones futuras para este modelo.

## 4.3. Modelo

Se considera una población de  $N$  individuos cuya opinión  $O$  emerge de una postura interna  $P$ , respaldada por un conjunto de argumentos. Los individuos participan en interacciones de a pares en las que intercambian argumentos, lo que eventualmente puede llevar a un cambio en sus opiniones. El estado de los agentes y la dinámica de las interacciones se detallan a continuación.

### 4.3.1. Estado de los Agentes

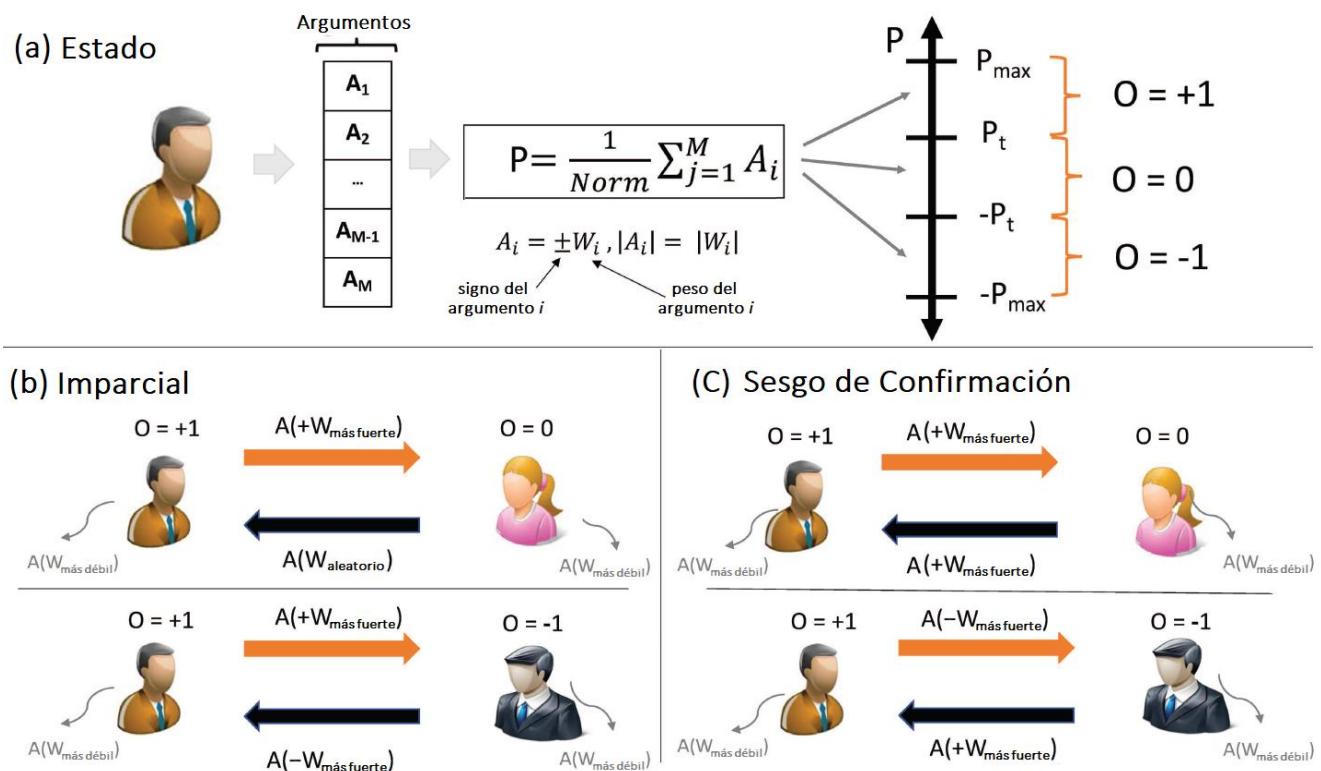
El estado de cada agente está representado por dos variables –no independientes– (Balenzuela et al., 2015) y un conjunto de argumentos que definen estas variables, y se esboza en el panel (a) de la Figura 1. La variable discreta  $O$  representa la opinión (pública) y toma tres valores posibles:  $O = 0, \pm 1$ . La variable continua  $P$  representa la postura del agente y puede tomar cualquier valor en el rango  $[-P_{\max}, P_{\max}]$ , mientras que el umbral  $P_t$  determina la opinión de un agente en función de  $P$ . Si la postura de un agente está en el rango  $[-P_{\max}, -P_t]$ , la opinión es  $O = -1$ ; si la postura de un agente está entre  $[-P_t, P_t]$ , la opinión es  $O = 0$ ; y dentro del rango  $[P_t, P_{\max}]$ , la opinión es  $O = 1$ . Se establece que  $P_{\max} = 3$  y  $P_t = 1$ , lo que asegura que los intervalos de postura que generan cada opinión sean iguales en tamaño. Se dirá que un agente está *orientado* si su opinión es  $O = \pm 1$  y *moderado* si  $O = 0$ .

De acuerdo con la Teoría de Argumentos Persuasivos, las posturas de los agentes ( $P$ ) se derivan de los argumentos que poseen en su memoria finita. Sea  $\mathbf{M}$  el número de argumentos que cada agente puede recordar y  $\mathbf{N}_A$  el número de argumentos disponibles para los agentes. Suponemos que la mitad de los  $N_A$  argumentos serán positivos (es decir, a favor del tema), la otra mitad serán negativos (es decir, en contra del tema). Además, cada argumento  $A$  tendrá un peso dado  $w$ , con valores enteros posibles de 1 a  $N_A/2$  y un signo  $V = \pm 1$ . De esta manera, para cada argumento  $A_j$

de una fuerza  $w_j$  y un signo  $V_j$  dados, existe un argumento opuesto de fuerza  $w_j$  y signo  $-V_j$ . La postura de un agente con  $M$  argumentos en memoria está definida por:

$$P = \frac{1}{N_A/2} \sum_{j=1}^M A_j = \frac{1}{N_A/2} \sum_{j=1}^M V_j \cdot w_j \quad [1]$$

donde el argumento  $A_j$  tiene peso  $w_j$  y signo  $V_j$ .



**Fig. 1. Estado de los Agentes y Dinámica de Interacción.** (a) Estado de los agentes. Cada agente posee un conjunto de  $M$  argumentos, que determinan su postura  $P$  según la Ec. (1). Si  $-P_{\max} \leq P < -P_t$  ó  $P_t < P \leq P_{\max}$ , el individuo mantendrá una opinión extrema ( $O = -1$  y  $O = +1$ , respectivamente). Si  $-P_t \leq P \leq P_t$ , los agentes mantendrán una opinión moderada ( $O = 0$ ) sobre el tema discutido. (b) Interacciones imparciales. En cada interacción bidireccional, los agentes orientados siempre compartirán su argumento más importante (es decir, el más fuerte), mientras que los agentes moderados compartirán uno al azar. (c) Interacciones sesgadas. En cada interacción, los agentes orientados incorporarán el argumento más fuerte de su oponente, alineado con su propia orientación (es decir, uno positivo si tienen  $O = +1$ ).

### 4.3.2. Dinámica

En este trabajo se consideran únicamente interacciones bidireccionales de a pares entre agentes.

Estas interacciones están mediadas por el intercambio de argumentos, que conduce a cambios en la postura del agente (según la Ec. [1]), y eventualmente pueden provocar cambios en las opiniones (si los cambios de postura cruzan los umbrales  $P_t$ ). En cada interacción, cada agente compartirá un argumento con su compañero, y recibirá otro de él. En este intercambio, ambos descartarán el argumento más débil de sus memorias.

A continuación, se considerarán tres modelos de interacciones: *imparciales, con sesgo de confirmación, y con homofilia*, según se detalla a continuación, y se esboza en los paneles (b) y (c) de la Figura 1.

#### 1 – Interacciones Imparciales

La idea básica aquí es que los agentes con opiniones extremas ( $O = \pm 1$ ) compartirán el argumento más fuerte alineado con su propia opinión, mientras que un agente moderado ( $O = 0$ ) compartirá uno al azar. Las reglas detalladas son las siguientes:

1 - Un agente orientado  $j$  siempre compartirá un argumento del mismo signo  $V$  que su opinión  $O$ , e intentará intercambiar el argumento más fuerte (por ejemplo, si  $O_j = +1$ , el argumento dado será el más fuerte con signo  $+1$ ). Si el otro agente ya tiene ese argumento, entonces  $j$  intercambiará el segundo más fuerte, y así sucesivamente. Esta regla es independiente de la opinión del otro agente.

2 - Un agente moderado siempre compartirá un argumento seleccionado al azar, independientemente de su signo. Como antes, si el otro agente ya tiene ese argumento, se elegirá otro al azar, independientemente de la opinión del otro agente.

Aquí, los agentes orientados con opiniones opuestas se comportan de la misma manera que si estuvieran interactuando con agentes que tienen la misma opinión. Sin embargo, y en contraposición a esta regla, una gran cantidad de evidencia empírica sugiere que las personas se comportan de manera sesgada cuando se enfrentan a argumentos que niegan sus creencias previas. Más específicamente, los humanos tienden a tener **sesgo de confirmación** (Parmley, 2006; Lehner et al., 2008; Nelson, 2014), según el cual las personas buscan abiertamente información que favorece sus creencias, expectativas y suposiciones preeexistentes (Koriat, 1980).

Este fenómeno implica descontar, o incluso ignorar, evidencia contraria, así como centrarse de manera desproporcionada en información favorable. Los sesgos de confirmación se han observado en una variedad de situaciones relevantes (por ejemplo, investigaciones criminales (O'Brien, 2009; Hill et al., 2008), deliberaciones médicas (Parmley, 2006; Tschan et al., 2009), investigación científica y prueba de hipótesis (Lehner et al., 2008; Lehner et al., 2009; Mynatt, 1977), difusión de noticias falsas, rumores o teorías de conspiración (Del Vicario, 2016), e incluso en la prevalencia de estereotipos sociales y discriminación consecuente (Nelson, 2014), y ha habido esfuerzos recientes para modelar su impacto en los estados colectivos (Del Vicario, 2016; Del Vicario et al., 2017; Starnini et al., 2016).

## **2 – Interacciones con Sesgo de Confirmación**

Para simular el sesgo de confirmación, se modifica ligeramente la regla de interacción entre agentes con opiniones opuestas: cada vez que los agentes interactúan, cada uno recibirá un

argumento favorable del otro, ignorando los argumentos no deseados. En otras palabras, el agente  $i$  con  $O_i = +1$  solo aceptará el argumento más fuerte de signo  $+1$  del agente  $j$  con  $O_j = -1$ . A su vez, el agente  $j$  recibirá el argumento más fuerte de signo  $-1$  del agente  $i$ . Todas las demás reglas permanecen sin cambios.

### 3 – Interacciones con Homofilia

Partiendo de una extensa literatura que sugiere que la similitud fomenta la atracción interpersonal, también se explora el papel de la homofilia en las interacciones de a pares entre agentes. Se define la similitud (Mas & Flache, 2013) entre los agentes  $i$  y  $j$  como  $S_{i,j} = 1 - \frac{|P_i - P_j|}{2P_{max}}$   $\forall j$ . Esta definición asegura que los agentes con la misma postura tendrán  $S_{i,j} = 1$ , y los agentes con posturas máximamente opuestas ( $P_i = -P_j$ , y  $|P_i| = P_{max}$ ) tendrán  $S_{i,j} = 0$ . La homofilia sugiere que cuanto más similares sean los agentes, mayor será la probabilidad de interacción que tengan. Esta probabilidad de interacción está regida por

$$Q_{i,j} = \frac{(S_{i,j})^h}{\sum_{p=1, p \neq i} (S_{i,p})^h} \quad [2]$$

donde  $h$  es un parámetro libre que representa la fuerza de la homofilia. En resumen, cuando dos agentes interactúan en esta condición, primero se selecciona al agente  $i$  con probabilidad uniforme, y luego al agente  $j$  con probabilidad  $Q_{i,j}$  de acuerdo con la Ecuación [2].

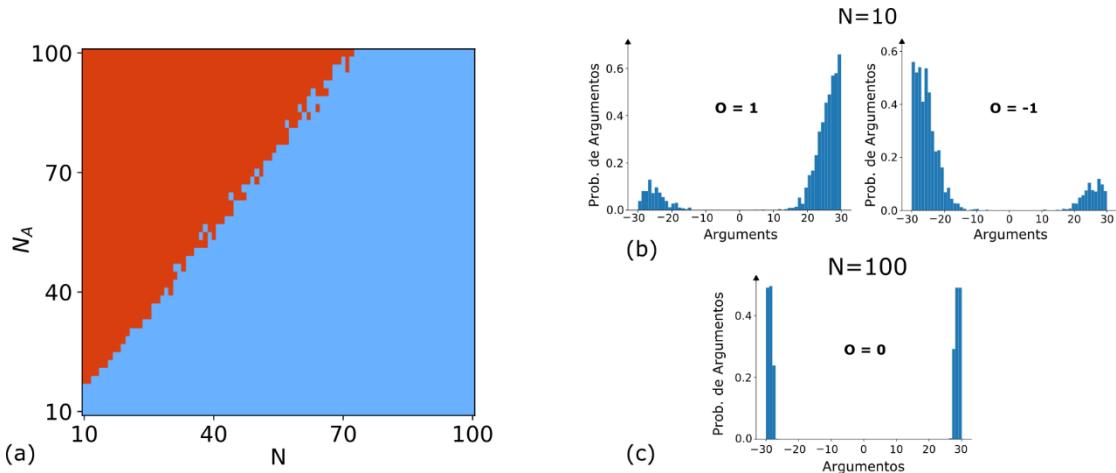
## 4.4. Resultados

Se considera una población de **N** agentes sin ninguna topología de conectividad subyacente. En  $t = 0$ , a cada agente se le proporcionan **M** argumentos diferentes, seleccionados al azar de la lista completa de **N<sub>A</sub>** argumentos disponibles. En cada paso de tiempo  $\Delta t = 1/N$ , se eligen al azar dos agentes, y se les hace interactuar siguiendo las reglas descritas anteriormente (un paso de tiempo corresponde a N interacciones aleatorias entre agentes). El sistema evoluciona hasta alcanzar un estado estacionario.

### 4.4.1. Modelo Imparcial

Se exploran los estados colectivos finales alcanzados por el sistema en función de tres parámetros libres del modelo, a saber, el número de agentes (**N**), el número de argumentos existentes (**N<sub>A</sub>**) y el tamaño de la memoria (**M**). Se inicia con un conjunto de  $M = 6$  argumentos fijos. La Figura 2(a) muestra el diagrama de fases con regiones espaciales donde predomina cada tipo de solución (es decir, donde es más probable cada una). Para definir las regiones, se estima la probabilidad de observar un estado dado s como  $N_s/N_{\text{tot}}$ , donde  $N_s$  es el número de realizaciones que convergieron a ese estado, y  $N_{\text{tot}} = 1000$  es el número total de simulaciones realizadas en ese punto del espacio de parámetros.

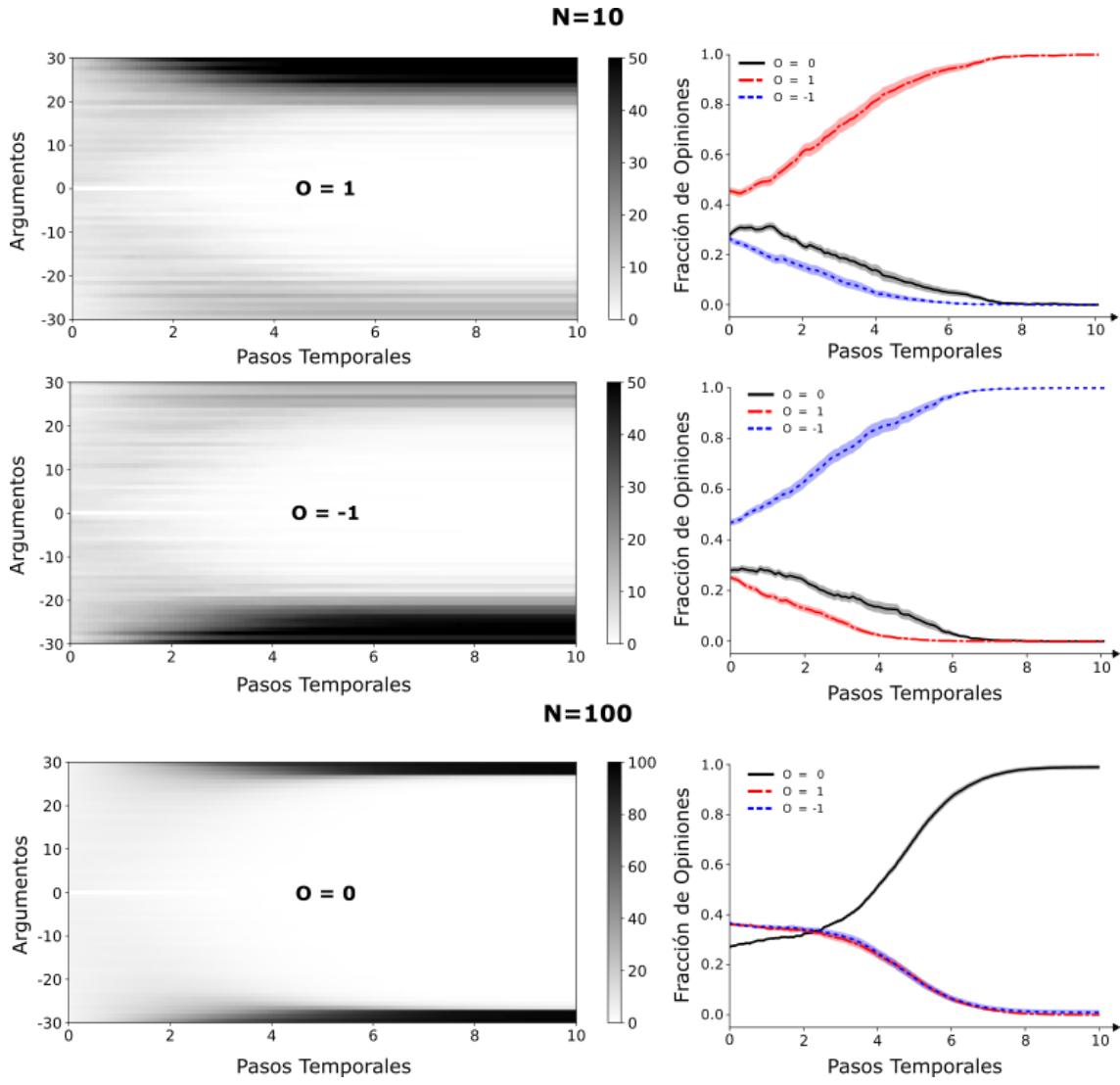
Se encuentran dos regiones bien separadas: una con consenso moderado (azul claro,  $O = 0$ ) y una región más pequeña con consenso orientado (rojo,  $O = \pm 1$ ). Por un lado, el consenso orientado es una solución dominante (aunque no la única), cuando hay aproximadamente más argumentos que agentes (la región superior izquierda en el panel (a) de la Figura 2). En esta región, la probabilidad de que todos los argumentos estén representados en el estado inicial del sistema se reduce, y, por lo tanto, aparecen casos particulares en los que los argumentos de un



**Fig. 2. Solución dominante en función de  $N$  y  $N_A$  para el modelo imparcial.** (a) Diagrama de fases del espacio de parámetros donde un estado dado es más probable que el otro. La región donde predomina el consenso moderado (abajo a la derecha) se muestra en azul claro, y la región donde predomina el consenso orientado (arriba a la izquierda) se muestra en rojo. (b) Distribución final de argumentos en el consenso orientado para  $N = 10$  y  $N_A = 60$ . (c) Distribución final de argumentos en el consenso moderado para  $N = 100$  y  $N_A = 60$ .

signo dado superan a los argumentos del signo opuesto, generando estados finales de consenso orientado.

Por otro lado, en la región donde hay menos argumentos que agentes, se observa que el consenso moderado está compuesto por agentes que tienen argumentos fuertes de ambos signos, como se muestra en el panel (c) de la Figura 2. Esto se debe a que la probabilidad de que todos los argumentos estén presentes al principio de las simulaciones es muy alta, lo que lleva a que sobrevivan  $M/2$  de los argumentos más fuertes de cada signo, alcanzándose así, inevitablemente, estados finales de consenso moderado. Ambos comportamientos se pueden observar en la evolución temporal de los argumentos y opiniones, como se muestra en la Figura 3. Específicamente, la evolución de los argumentos en la región moderada ( $N = 100$ , tercera fila)



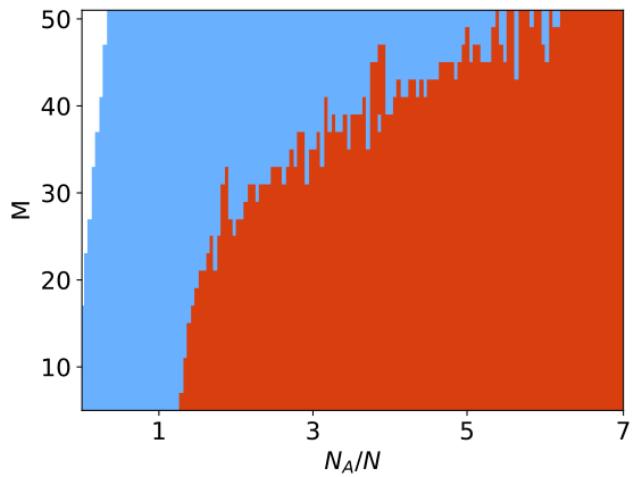
**Fig. 3. Dinámica del modelo imparcial.** Cada uno de los seis paneles se construyó en base a 100 realizaciones aleatorias de las condiciones iniciales ( $N_{\text{tot}} = 100$ ), con parámetros  $N_A = 60$  argumentos, y  $M = 6$  tamaño de memoria. Las dos primeras filas corresponden a la región de consenso orientado (la región roja en la Fig. 2,  $N = 10$  agentes), y la tercera fila corresponde a la región de consenso moderado (la región azul claro en la Fig. 2,  $N = 100$ ). Un paso de tiempo corresponde a  $N$  interacciones. **Columna izquierda:** histogramas que muestran la evolución de los argumentos para diferentes estados finales del sistema. Se saturó la escala para una mejor apreciación de los argumentos iniciales. **Columna derecha:** evolución de la fracción de agentes en cada estado de opinión. Las regiones sombreadas denotan los intervalos de confianza del 95% respectivos.

muestra claramente cómo todos los argumentos están inicialmente presentes, y solo los más fuertes de cada signo permanecen en el estado final. Además, en la región orientada ( $N = 10$ , dos primeras filas), no todos los argumentos están inicialmente presentes (como se evidencia por la heterogeneidad de los estados iniciales) y, aunque algunos argumentos fuertes pueden permanecer para la opinión opuesta, la dominante tiene argumentos más fuertes en promedio. En la misma figura se puede observar la evolución de la dinámica de las opiniones para las mismas regiones.

Dado el comportamiento observado en la Figura 2(a), donde la dominancia de cada solución depende aproximadamente de la relación  $N_A/N$ , en la Figura 4 se traza el estado colectivo dominante en función de  $M$  y de  $N_A/N$ . Se observan aquí las mismas soluciones que en el gráfico anterior: consenso orientado (rojo) y consenso moderado (azul). Podemos ver que los estados de consenso orientado dominan el comportamiento colectivo para valores grandes de  $N_A/N$  (lo que significa que hay muchos más argumentos que agentes), especialmente si el tamaño de la memoria no es muy grande. Esto se puede explicar de manera similar: el tamaño de la memoria está relacionado con el número de argumentos disponibles en las condiciones iniciales. Cuando el número de agentes es pequeño en relación con el número de argumentos, para un tamaño de memoria reducido, muchos argumentos faltarán al principio de la simulación, lo que llevará al consenso orientado. Sin embargo, cuando el tamaño de la memoria aumenta y/o el número de argumentos disminuye en relación con el número de agentes (valores bajos de  $N_A/N$ ), es más probable que todos los argumentos existentes estén presentes, lo que lleva al consenso moderado. Solo para valores grandes de  $N_A/N$  y valores pequeños de memoria, se vuelve posible encontrar consenso orientado.

#### 4.4.2. Modelo con Sesgo de Confirmación

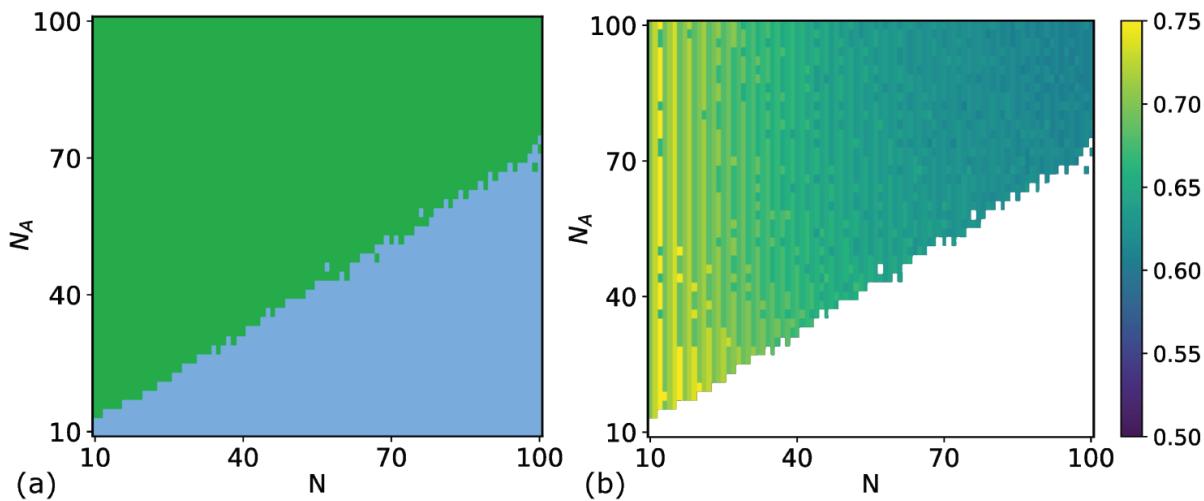
En esta sección se expone el papel desempeñado por el sesgo de confirmación. Se repitió la exploración de parámetros, pero centrándose una vez más en la relación entre  $N$  y  $N_A$ , para  $M = 6$  fijo (Figura 5(a)).



**Fig. 4. Solución dominante en función de  $M$  y  $N_A/N$  para el modelo imparcial.** Azul claro: región donde es más probable el consenso moderado (arriba a la izquierda). Rojo: Región donde predomina el consenso orientado (abajo a la derecha). La región blanca corresponde a los casos donde  $M \geq N_A$ , donde no se olvidan argumentos (lo cual es físicamente irrelevante). Utilizamos 1000 realizaciones aleatorias de las condiciones iniciales para cada par de parámetros ( $N_{\text{tot}} = 1000$ ). Se barrió  $N$  y  $N_A$  de 10 a 100, el primero en incrementos de 1, y el último en incrementos de 2. Un valor dado de  $N_A/N$  proviene del promedio de todas las combinaciones posibles de  $N$  y  $N_A$  que dan ese valor específico.

Se encontró que las soluciones con consenso orientado han desaparecido del espacio de parámetros, y han sido reemplazadas por estados de "bipolarización", donde dos opiniones orientadas coexisten en la población. Esta solución ahora se ha vuelto prevalente en el espacio de parámetros, siendo el estado más probable en todas las regiones previamente dominadas por

estados orientados, y también reduciendo el tamaño de la región de consenso moderado. Al igual que en el modelo no sesgado, el consenso moderado se encontró principalmente para  $N > N_A$ , lo que también se explica por la presencia de todos los argumentos existentes en el estado inicial del sistema.



**Fig. 5. Solución dominante en función de  $N$  y  $N_A$  para el modelo de sesgo de confirmación.** (a) Regiones del espacio de parámetros donde un estado dado es más probable que los demás. Azul claro: región donde es más probable el consenso moderado (abajo a la derecha). Verde: región donde predomina la bipolarización (arriba a la izquierda). (b) Número medio de agentes con una opinión orientada dominante, normalizado por  $N$ . En ambos paneles, cada punto es un promedio sobre 1000 realizaciones aleatorias de las condiciones iniciales.

La razón subyacente a la aparición de la bipolarización está relacionada con la nueva regla de interacción propuesta en el modelo para tener en cuenta el sesgo de confirmación. Ahora, cuando dos agentes con opiniones opuestas interactúan, intercambian argumentos que refuerzan sus propias posiciones. Esto actúa efectivamente como una fuerza de repulsión, que

impulsa a los agentes orientados hacia posturas extremas. Se debe tener en cuenta que la bipolarización no es necesariamente equilibrada: puede haber más agentes de una opinión que de la otra. Esto se puede observar en la Figura 5(b). Para valores más pequeños de N, es posible encontrar más agentes de una opinión que de la otra (de color verde claro a amarillo). Para valores más grandes de N, la bipolarización tiende a volverse equilibrada (color verde oscuro).

La exploración del parámetro M arroja resultados similares, solo que esta vez la bipolarización adopta el papel ocupado anteriormente por el consenso orientado, que a su vez desaparece por completo para valores de M en el rango medio (se omite el gráfico para evitar redundancias).

Los resultados obtenidos, tanto para modelos sesgados como no sesgados, motivan la exploración de la transición entre estos dos casos. Con este fin, se definió un nuevo parámetro: la **probabilidad de sesgo de confirmación**,  $P_{CB}$ . Cada vez que dos agentes orientados con opiniones opuestas interactúan,  $P_{CB}$  determina la probabilidad de hacerlo con sesgo de confirmación. Es importante tener en cuenta que el modelo no sesgado es equivalente a  $P_{CB} = 0$ , y el modelo sesgado es equivalente a  $P_{CB} = 1$ . Esto lleva a esperar que los estados finales del sistema sean más similares al modelo no sesgado para valores cercanos a 0, y más similares al modelo sesgado para valores cercanos a 1. Sin embargo, el resultado fue inesperado: no hubo estados de bipolarización para ningún valor de  $P_{CB} < 1$  (ver Figura 6(a)). Además, los consensos moderado y orientado permanecieron, en su mayoría, sin cambios.

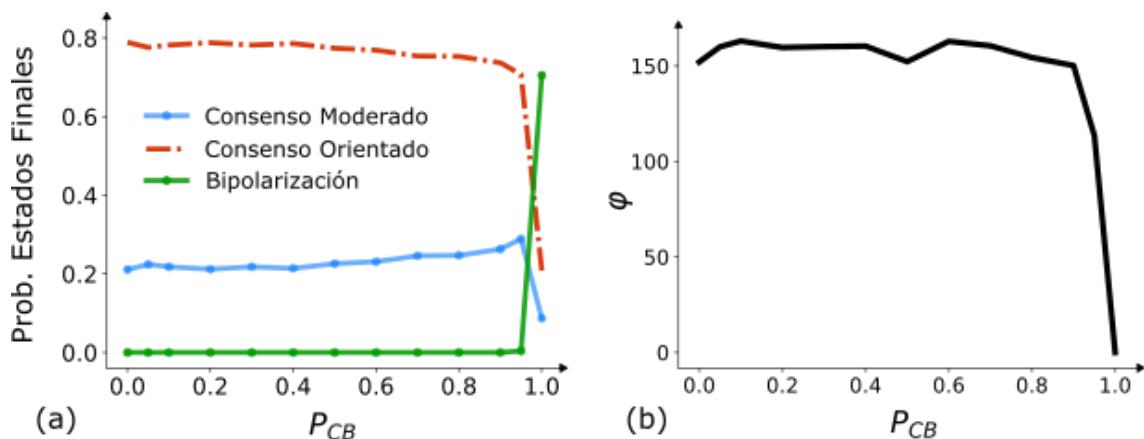
Para estudiar más a fondo este fenómeno, se definió una nueva magnitud: el flujo de información contraria,  $\varphi$ , que mide los pesos acumulados de los argumentos de opinión opuesta adquiridos por un agente orientado. Esto puede ocurrir ya sea a través de interacciones con un

agente moderado (de quien toma un argumento elegido al azar) o a través de interacciones con otro agente orientado, pero con una opinión opuesta.

$$\varphi = \sum_{int} \varphi_{int} = \sum_{int} (w_{i,int} - w_{j,int}) \quad [3]$$

Para cada interacción entre agentes orientados con opiniones opuestas, se sumaron los pesos de los argumentos que han sido intercambiados. De esta manera, se obtiene el flujo de información contraria para cada interacción,  $\varphi_{int}$ . Sumando todas las interacciones, se llega al flujo total de información contraria,  $\varphi$  (Ec. (3)). Los resultados de este flujo de información se muestran en la Figura 6(b).

La Figura 6 muestra que sólo se encuentra bipolarización para  $P_{CB} = 1$ , y esto, a su vez, está relacionado con  $\varphi = 0$ . Esto demuestra que la bipolarización se alcanza debido a la falta de información contraria adquirida a través de la interacción entre agentes. Para cualquier otro valor de  $P_{CB}$ ,  $\varphi \neq 0$ , lo que significa que, eventualmente, los agentes orientados obtienen argumentos contrarios, volviéndose moderados.



**Fig. 6. Transición entre modelos no sesgados y sesgados.** (a) Probabilidad de cada estado final vs  $P_{CB}$ . (b)  $\varphi$  vs  $P_{CB}$ . Valores numéricos utilizados:  $N_{tot} = 1000$ ,  $N = 10$ ,  $N_A = 60$  y  $M = 6$ .

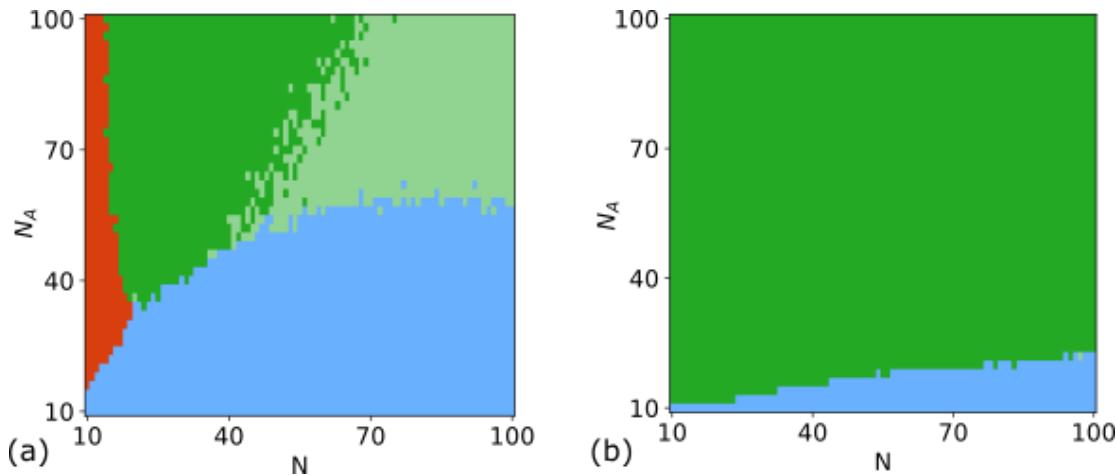
#### **4.4.3. Modelo con Homofilia**

Investigaciones previas han demostrado que la combinación de homofilia y el intercambio de argumentos puede producir estados de bipolarización (Mas & Flache, 2013). Para explorar el papel desempeñado por la homofilia en la polarización, se enriqueció el modelo incluyendo su dinámica. La Figura 7 presenta los resultados tanto para el modelo imparcial con homofilia (a) como para el modelo sesgado con homofilia (b), de acuerdo con la Ec. (2), para  $h = 4$ . Como antes, la preponderancia del consenso moderado se representa en azul claro, el consenso orientado en rojo, y la bipolarización en verde, y se agregó una región verde claro que corresponde a la preponderancia de estados metaestables de bipolarización. Estos últimos podrían conducir a estados finales de consenso moderado, consenso orientado o bipolarización. Estos estados metaestables apenas se encuentran presentes para valores bajos de homofilia; sin embargo, para valores más altos ( $h > 3$ ), se vuelven predominantes en una región del espacio. Solo se pueden encontrar cuando no está presente el sesgo de confirmación. El número de pasos de tiempo necesario para que pierdan estabilidad, y alcancen un estado estable final, es mucho mayor que el número medio requerido por los otros estados (más de dos desviaciones estándar).

Una conclusión importante extraída de la Figura 7 es que, sin sesgo de confirmación, la bipolarización solo puede surgir en presencia de homofilia. Para diferentes valores de  $h$ , se observa que la región de bipolarización se agranda con  $h$  (se omiten los gráficos para otros valores para evitar redundancias). El consenso orientado es prácticamente reemplazado por la bipolarización. Además, cuando tanto el sesgo de confirmación como la homofilia están presentes (Fig. 7(b)), la bipolarización domina casi todo el diagrama de fases, salvo por una

pequeña región de consenso moderado. Esto sugiere que la homofilia y el sesgo de confirmación se refuerzan mutuamente, produciendo aún más bipolarización.

La razón detrás de los estados de bipolarización producidos por la homofilia es que los agentes que tienen valores extremos opuestos de postura (es decir, algunos con  $P = -3$  y otros con  $P = 3$ ) no pueden interactuar (ver Eq. (2)). Si la probabilidad de interacción entre agentes con postura similar es mucho mayor que la de aquellos con posturas diferentes, entonces las interacciones



**Fig. 7. Efectos de la homofilia en modelos sesgados e imparciales.** Regiones del espacio de parámetros donde un estado dado es más probable que otro, para el modelo imparcial **(a)**, y para el modelo con sesgo de confirmación **(b)**. En ambos casos,  $h = 4$ , y se utilizaron 1000 realizaciones aleatorias de las condiciones iniciales. Azul claro: región donde es más probable el consenso moderado que los demás (región inferior). Rojo: región donde predomina el consenso orientado (región izquierda). Verde: región donde predomina la bipolarización. Verde claro: región donde los estados metastables de bipolarización son los estados más probables.

tienden a producir grupos de agentes orientados en los lados opuestos del espectro. Después de algunas interacciones sucesivas dentro del grupo, todos los agentes terminan con todos los argumentos más fuertes de su propio lado y, por lo tanto, con los valores de postura más extremos. La combinación conjunta de homofilia y sesgo de confirmación hace que estos efectos sean más fuertes.

Con todo esto en mente, se puede concluir que tanto la homofilia como el sesgo de confirmación producen estados de bipolarización en nuestro modelo. Es posible observar que una homofilia fuerte juega el mismo papel que el sesgo de confirmación: obstruir el flujo de información contraria entre grupos con opiniones opuestas. Cuando el sesgo de confirmación está presente, no hay flujo debido a la falta de interacciones entre agentes con opiniones opuestas; mientras tanto, cuando predomina la homofilia, es debido a un proceso de refuerzo en las interacciones, como se mostró anteriormente.

## 4.5. Discusión General

En este capítulo se presentó un modelo donde el cambio de opinión está mediado por el intercambio de argumentos, y se analizaron los estados colectivos en términos de reglas dinámicas simples. Una interpretación posible de estas reglas de interacción es que dos agentes participan en una discusión y escuchan los argumentos que el otro tiene para ofrecer. En una discusión imparcial, un agente con una opinión orientada tiende a convencer a cualquier otro de su postura, lo que lleva al otro agente a incorporar un argumento de esa opinión (el más fuerte, en este caso). Sin embargo, si la discusión está sesgada, cuando interactúan dos agentes con opiniones opuestas, solo escucharán lo que favorece a las opiniones que ya tienen, e ignorarán

los argumentos opuestos. Cualquier "concesión" será exacerbada por el otro agente y, por lo tanto, los agentes podrían terminar aún más convencidos de su postura previa después de discutir con alguien de postura opuesta. Este fenómeno, conocido como el "efecto boomerang", donde una persona intenta persuadir a otros y refuerza involuntariamente sus creencias previas, se ha observado previamente en investigaciones empíricas (Cohen, 1962). En contraste, los agentes orientados pueden volverse más moderados al interactuar con agentes moderados, pero también podrían volverse más orientados (y esto no cambia en ambas variantes del modelo). Nuevamente, este efecto, donde las opiniones se vuelven más extremas después de interactuar con otras personas, se ha observado extensamente en la investigación psicológica (Moscovici & Zavalloni, 1969).

El modelo imparcial presentado tuvo éxito en la reproducción de estados de consenso, ya sea un consenso orientado ( $O = 1$  o  $O = -1$  para todos los agentes) o un consenso moderado ( $O = 0$ ). La región en el espacio de parámetros donde predomina este último es más grande, y en su mayoría se limita a la región donde el número de argumentos es menor que el número de agentes ( $N_A < N$ ), lo que significa que hay una alta probabilidad de que todos los argumentos estén presentes al principio de las simulaciones. Si bien los cambios en el tamaño de la memoria ( $M$ ) son comparativamente menos importantes, el aumento en el número de argumentos lleva a un aumento en los estados de consenso orientado. El parámetro  $P_t$ , que es el umbral entre opiniones, no desempeña un papel relevante, siempre y cuando las tres opiniones se asignen a intervalos de postura de la misma longitud (Balenzuela et al., 2015). Sin embargo, un análisis más profundo que involucre una distribución de umbrales en lugar de un valor fijo podría acarrear implicancias interesantes, algo que queda abierto para futuras investigaciones.

El estado de consenso moderado merece un análisis cuidadoso. Esto se debe a que, en este modelo, a priori, los agentes podrían tener una opinión moderada por dos razones: (1) porque poseen argumentos débiles o (2) porque poseen argumentos fuertes de opiniones opuestas, de modo que ambos tipos de argumentos se cancelan (según la Ec. (1)), y su postura se acerca a cero. Dada la dinámica de nuestro modelo, solo puede producirse polarización de argumentos y, por lo tanto, solo es posible observar el segundo escenario, donde los agentes moderados son el resultado de agentes que tienen argumentos fuertes de ambos lados de la discusión.

Se podría argumentar que, suponiendo que la fuerza de los argumentos puede medirse objetivamente, la discusión sin sesgos inevitablemente terminará en consenso (ya sea moderado u orientado): si los argumentos se comparten libremente y se aceptan, tarde o temprano los argumentos más fuertes predominarán, mientras que el resto se olvidará, y todos tendrán esos mismos argumentos. Sin embargo, la presencia de un sesgo de confirmación restringe el flujo libre de información: alguien con cierta opinión incorporará argumentos favorables, pero subestimará o ignorará la información contraria. Esto, a su vez, podría producir bipolarización, lo que llevó a considerar una variación del modelo imparcial, con la inclusión de un sesgo de confirmación.

Al introducir el sesgo de confirmación, aparecen estados de bipolarización. A diferencia del caso anterior, la bipolarización reemplaza al consenso orientado, y su región de preponderancia se vuelve más grande que la del estado de consenso moderado. Sin embargo, después de introducir una probabilidad de sesgo de confirmación ( $P_{CB}$ ), no fue posible encontrar una forma concreta de generar una transición suave entre el modelo sesgado y el imparcial. En cambio, se encontró que la bipolarización sólo estaba presente cuando  $P_{CB} = 1$ , lo que podría explicarse por el flujo de

información contraria ( $\varphi$ ). Cuando  $\varphi = 0$ , no se puede transmitir información contraria, y la bipolarización se vuelve estable. En cualquier otro caso, la bipolarización termina llevando al consenso.

La introducción de la homofilia indujo la aparición de bipolarización en ausencia de sesgo de confirmación. Para valores altos del parámetro de homofilia ( $h$ ), la bipolarización prácticamente dominaba todas las regiones donde el consenso orientado solía ser el estado más probable, y algunas partes de la región donde el consenso moderado era preponderante. También es posible encontrar una región donde los estados metastables de bipolarización son predominantes. Adicionalmente, cuando se incluía el sesgo de confirmación, la bipolarización dominaba la mayor parte del espacio de parámetros (con una región reducida de consenso moderado para valores pequeños de  $N_A$ ). Así, la superposición de ambos mecanismos produce aún más estados de bipolarización que con estos mecanismos por separado.

Dada la ubicuidad del sesgo de confirmación y la homofilia en las interacciones sociales, el modelo de intercambio de argumentos desarrollado en este capítulo podría ser muy útil para informar futuros experimentos relacionados con la dinámica de formación de opinión, y podría explicar potencialmente los estados colectivos macroscópicos observados en las sociedades. La versatilidad de las reglas de interacción permite modificaciones simples –por ejemplo, la introducción de redes de interacción con diferentes topologías. Estas podrían proporcionar nuevas ideas sobre cómo las personas interactúan, modifican sus opiniones, y toman decisiones importantes en diferentes escalas, desde reuniones de trabajo reducidas hasta elecciones presidenciales.

## Capítulo 5

### 5. Discusión General

La toma de decisiones grupales constituye un área de investigación abierta de gran relevancia para diversas áreas del conocimiento, como la Psicología Experimental, la Neurociencia Cognitiva, y la Física Estadística. Existen diferentes abordajes, estrategias y objetivos que intervienen en su estudio, como, por ejemplo, identificar métodos óptimos para la obtención de conocimiento y resolución de problemas de una multitud; resolver problemas grupales por medio de estrategias eficientes de interacción; o la generación de consenso en problemas para los que no necesariamente existe una respuesta correcta. En todos los casos, se trata de problemas muy relevantes para las decisiones democráticas. En esta tesis se presentaron tres estrategias diseñadas para mejorar las chances de alcanzar consensos grupales, y extraer de estos información útil para la resolución de problemas.

En el capítulo 2 se expuso una estrategia novedosa para mejorar la sabiduría de las multitudes (Surowiecki, 2005), con implicancias tanto prácticas como teóricas. Mediante el empleo del efecto de anclaje (Tversky y Kahneman, 1974), se demostró que es posible reducir el error colectivo aumentando la diversidad de opiniones de la multitud (Page, 2007). En cuatro experimentos comportamentales se observó que, independientemente de las diferencias en el tamaño de la muestra, el país de implementación, el uso de cantidades acotadas o no acotadas, y si la tarea involucraba estimación o pronóstico, las multitudes extremizadas siempre producían estimaciones con un error colectivo más bajo, a pesar de presentar un mayor error individual. Este abordaje novedoso podría inspirar futuras investigaciones destinadas a aumentar la

sabiduría de las multitudes mediante una estrategia similar, potencialmente combinada con otras estrategias de reducción de sesgo que han demostrado ser útiles para este fin.

En el capítulo 3 se estudiaron en detalle los mecanismos y estrategias detrás de estimaciones colectivas realizadas por grupos reducidos de personas. Se tomó como punto de partida un estudio previo (Navajas et al., 2018) que sugería que la deliberación grupal para llegar a un valor consensuado, seguida de una agregación de estimativos provenientes de varios grupos, es una estrategia extremadamente efectiva para mejorar la sabiduría de las multitudes clásica. En esta línea se estudió qué estrategias seguían estos grupos para llegar a tales consensos, algo que permanecía inexplorado hasta ahora. Los resultados presentados sugieren que esta mejora en la exactitud no es simplemente el resultado de agrupar estimaciones individuales iniciales, sino que es facilitada por un proceso de razonamiento y cálculo colectivo dentro de los grupos (similar al Método de Fermi). En particular, los grupos que emplean predominantemente una estrategia similar al Método de Fermi exhiben tasas de error significativamente más bajas en comparación con aquellos que emplean estrategias alternativas, como simplemente agregar estimaciones individuales. Esto destaca la eficacia potencial de instruir a los participantes para que utilicen técnicas cognitivas específicas, optimizando así los procesos de toma de decisiones grupales. En línea con esto, los hallazgos presentados permiten dilucidar la compleja interacción entre la influencia social y la toma de decisiones colectivas, yendo más allá de la indagación académica, y ofreciendo ideas prácticas para mejorar los procesos de toma de decisiones en ámbitos diversos.

Por último, en el capítulo 4 se exploró el proceso de formación de consenso y bipolarización mediante un modelo de agentes. Esto profundiza el análisis de los capítulos anteriores (con problemas fácticos), expandiendo el dominio hacia problemas de opinión, que no

necesariamente presentan una respuesta correcta. Se presentó un modelo donde el cambio de opinión está mediado por el intercambio de argumentos, y se analizaron los estados colectivos en términos de reglas dinámicas simples. Este modelo, que representaba interacciones individuales microscópicas con un intercambio de argumentos que podía ser sesgado (con sesgo de confirmación y/o homofilia de interacción) o no, podría ser muy útil para informar futuros experimentos relacionados con la dinámica de formación de opinión, y podría explicar potencialmente los estados colectivos macroscópicos observados en las sociedades (como la “grieta”, el fenómeno de polarización que afecta a muchas de ellas en el mundo moderno). La versatilidad de las reglas de interacción del modelo permite modificaciones simples que podrían proporcionar nuevas ideas sobre cómo las personas interactúan, modifican sus opiniones, y toman decisiones importantes en diferentes escalas, desde reuniones de trabajo reducidas hasta elecciones presidenciales.

El primer estudio sobre la sabiduría de las multitudes (Galton, 1907) fue considerado como una demostración empírica de que los principios de agregación democrática pueden ser confiables y eficientes. Desde entonces, ha habido numerosos avances en el estudio de qué factores contribuyen a formar multitudes más sabias, grupos más eficientes, y respuestas consensuadas tanto en problemas intelectivos como no intelectivos. Hoy en día, cuando las opiniones políticas tienden a extremizarse, y la polarización adquiere máxima relevancia, entender estos interrogantes se torna especialmente importante, tanto para fines teóricos (como el estudio del valor epistémico de la democracia) como prácticos (mejorar las discusiones y alcanzar consensos).

## 6. Bibliografía

- Abramowitz, A. (2010). The disappearing center: Engaged citizens, polarization, and American democracy. Yale University Press.
- Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science advances*, 7(36). <https://doi.org/10.1126/sciadv.abf4393>
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science*, 329(5995), 1081–1085. <https://doi.org/10.1126/science.1185718>
- Balenzuela, P., Pinasco, J. P., & Semeshenko, V. (2015). The undecided have the key: Interaction-driven opinion dynamics in a three state model. *PloS one*, 10(10), e0139572.
- Ball, P. (2004). Critical mass: How one thing leads to another. Macmillan.
- Baronchelli, A. (2018). The emergence of consensus: a primer. *Royal Society open science*, 5(2), 172189.
- Becker, J., Porter, E., & Centola, D. (2019). The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 116(22), 10717-10722.
- Berger, T. (2001). Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural economics*, 25(2-3), 245-260.
- Billari and A. Prskawetz, Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour, Springer Science & Business Media (2012).
- Billari, F. C., & Prskawetz, A. (Eds.). (2003). Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour. Springer Science & Business Media.

Bohman, J., & Rehg, W. (1997). Deliberative democracy: Essays on reason and politics. MIT press.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.

Burnstein, E., & Vinokur, A. (1973). Testing two classes of theories about group induced shifts in individual choice. *Journal of experimental social psychology*, 9(2), 123-137.

Burnstein, E., & Vinokur, A. (1975). What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology*, 11(5), 412-426.

Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of experimental social psychology*, 13(4), 315-332.

Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The journal of abnormal and social psychology*, 62(3), 713.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of modern physics*, 81(2), 591.

Chmiel, A., Sienkiewicz, J., & Sznajd-Weron, K. (2017). Tricriticality in the q-neighbor Ising model on a partially duplex clique. *Physical Review E*, 96(6), 062137.

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, 60(3), 581-588.

Cohen, A. R. (1962). A dissonance analysis of the boomerang effect. *Journal of Personality*, 30(1).

- Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of personality and social psychology*, 85(5), 808.
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., ... & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214, 325-346.
- Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., Bearden, C. E., & Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, 17(1), 67–75. <https://doi.org/10.1002/wps.20491>
- Cox, J. T., & Griffeath, D. (1986). Diffusive clustering in the two dimensional voter model. *The Annals of Probability*, 347-370.
- Crokidakis, N., Blanco, V. H., & Anteneodo, C. (2014). Impact of contrarians and intransigents in a kinetic model of opinion dynamics. *Physical Review E*, 89(1), 013310.
- Dandekar, P., Goel, A., & Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15), 5791-5796.
- De Condorcet, N. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. *L'impremerie royale*.
- Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04), 87-98.
- DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, 69(345), 118-121.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the national academy of Sciences*, 113(3), 554-559.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific reports*, 7(1), 40391.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685-686.

Filatova, T., Verburg, P. H., Parker, D. C., & Stannard, C. A. (2013). Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental modelling & software*, 45, 1-7.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Jasss-The journal of artificial societies and social simulation*, 20(4), 2.

Frachebourg, L., & Krapivsky, P. L. (1996). Exact results for kinetics of catalytic reactions. *Physical Review E*, 53(4), R3009.

Frey, V., & Van de Rijt, A. (2021). Social influence undermines the wisdom of the crowd in sequential decision making. *Management science*, 67(7), 4273-4286.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics*, 40(1), 35-42.

Galam, S. (1986). Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach. *Journal of Mathematical Psychology*, 30(4), 426-434.

Galam, S. (2002). Minority opinion spreading in random geometry. *The European Physical Journal B-Condensed Matter and Complex Systems*, 25, 403-406.

Galton, F. (1907). Vox populi. *Nature* 7, 450–451.

Goldenberg, A., Abruzzo, J. M., Huang, Z., Schöne, J., Bailey, D., Willer, R., ... & Gross, J. J. (2023). Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour*, 7(2), 219-230.

Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in cognitive sciences*, 9(9), 424-430.

Granovetter, M., & Soong, R. (1986). Threshold models of interpersonal effects in consumer demand. *Journal of Economic Behavior & Organization*, 7(1), 83-99.

Gürçay, B., Mellers, B. A., & Baron, J. (2015). The Power of Social Influence on Estimation Accuracy. *Journal of Behavioral Decision Making*, 28(3), 250–261.  
<https://doi.org/10.1002/bdm.1843>

Gutiérrez, L., & Keith, B. (2019). A Systematic Literature Review on Word Embeddings. In J. Mejia, M. Muñoz, Á. Rocha, A. Peña, & M. Pérez-Cisneros (Eds.), *Trends and Applications in Software Engineering* (Vol. 865, pp. 132–141). Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-01171-0\\_12](https://doi.org/10.1007/978-3-030-01171-0_12)

Haken, H. (1983). An introduction nonequilibrium phase transitions and self-organization in physics. Chemistry, and Biology, 206-216.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, 5(3).

Hegselmann, R., & Krause, U. (2005). Opinion dynamics driven by various ways of averaging. *Computational Economics*, 25, 381-405.

Helbing, D. (2010). Quantitative sociodynamics: stochastic methods and models of social interaction processes. Springer Science & Business Media.

Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and criminological psychology*, 13(2), 357-371.

Hinsz, V. B., & Davis, J. H. (1984). Persuasive arguments theory, group polarization, and choice shifts. *Personality and Social Psychology Bulletin*, 10(2), 260-268.

Holley, R. A., & Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, 643-663.

Hong, L., & Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385-16389.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual review of political science*, 22, 129-146.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161-1166.

Jager, W., & Amblard, F. (2005). Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational & Mathematical Organization Theory*, 10, 295-303.

Jayles, B., Kim, H. R., Escobedo, R., Cezera, S., Blanchet, A., Kameda, T., ... & Theraulaz, G. (2017).

How social information can improve estimation accuracy in human groups. *Proceedings of the National Academy of Sciences*, 114(47), 12620-12625.

Jönsson, M. L., Hahn, U., & Olsson, E. J. (2015). The kind of group you want to belong to: Effects of group structure on group accuracy. *Cognition*, 142, 191-204.

Juni, M. Z., & Eckstein, M. P. (2015). Flexible human collective wisdom. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1588.

Kameda, T., Toyokawa, W., & Tindale, R. S. (2022). Information aggregation and collective intelligence beyond the wisdom of crowds. *Nature Reviews Psychology*, 1(6), 345-357.

Kao, A. B., & Couzin, I. D. (2014). Decision accuracy in complex environments is often maximized by small group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 281(1784), 20133305.

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., ... & Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of The Royal Society Interface*, 15(141), 20180130.

Karachiwalla, R., & Pinkow, F. (2021). Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creativity and innovation management*, 30(3), 563-584.

Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., ... & Meyer, P. (2017). Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355(6327), 820-826.

Kohler, T. A., & Gumerman, G. G. (Eds.). (2000). Dynamics in human and primate societies: Agent-based modeling of social and spatial processes. Oxford University Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory*, 6(2), 107.

Krapivsky, P. L. (1992). Kinetics of a monomer-monomer model of heterogeneous catalysis. *Journal of Physics A: Mathematical and General*, 25(22), 5831.

Krizan, Z., & Baron, R. S. (2007). Group polarization and choice-dilemmas: how important is self-categorization? *European Journal of Social Psychology*, 37(1), 191-201.

Kurahashi-Nakamura, T., Mäs, M., & Lorenz, J. (2016). Robust clustering in generalized boundedconfidence models. *JASSS-The Journal of Artificial Societies and Social Simulation*, 19(4), 7.

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., Zalaudek, I., & Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, 113(31), 8777-8782.

Laughlin, P. R., & Earley, P. C. (1982). Social combination models, persuasive arguments theory, social comparison theory, and choice shift. *Journal of Personality and Social Psychology*, 42(2), 273.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... Van Alstyne, M. (2009). Social science: Computational social science. *Science*, 323(5915), 721–723.

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062.

Lehner, P. E., Adelman, L., Cheikes, B. A., & Brown, M. J. (2008). Confirmation bias in complex analyses. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3), 584-592.

Lehner, P. E., Adelman, L., DiStasio, R. J., Erie, M. C., Mittel, J. S., & Olson, S. L. (2008). Confirmation bias in the analysis of remote sensing data. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(1), 218-226.

Liggett, T. M. (2012). Interacting particle systems, Springer Science & Business Media, vol. 276 Berlin.

Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12), 1819-1838.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, 108(22), 9020-9025.

Lorenz, J., Neumann, M., & Schröder, T. (2021). Individual attitude change and societal dynamics: Computational experiments with psychological theories. *Psychological Review*, 128(4), 623.

Madirolas, G., & de Polavieja, G. G. (2015). Improving collective estimations using resistance to social influence. *PLoS computational biology*, 11(11), e1004594.

Majorana, E. (1942). Il valore delle leggi statistiche nella fisica e nelle scienze sociali. *Scientia*, 36(71).

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of personality and social psychology*, 107(2), 276.

- Martins, A. C. (2008). Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(04), 617-624.
- Mäs, M., & Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PLoS one*, 8(11), e74516.
- McGuire, T. W., Kiesler, S., & Siegel, J. (1987). Group and computer-mediated discussion effects in risk decision making. *Journal of personality and social psychology*, 52(5), 917.
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014a). Psychological Strategies for Winning a Geopolitical Forecasting Tournament. *Psychological Science*, 25(5), 1106–1115.  
<https://doi.org/10.1177/0956797614524255>
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014b). Psychological Strategies for Winning a Geopolitical Forecasting Tournament. *Psychological Science*, 25(5), 1106–1115.  
<https://doi.org/10.1177/0956797614524255>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2), 57-74.
- Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). *Advances in Pre-Training Distributed Word Representations* (arXiv:1712.09405). arXiv. <http://arxiv.org/abs/1712.09405>
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of personality and social psychology*, 12(2), 125.
- Mussweiler, T., & Strack, F. (2001). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, 19(2), 145-160.

- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29(1), 85-95.
- Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2), 126-132.
- Nelson, J. A. (2014). The power of stereotyping and confirmation bias to overwhelm accurate assessment: The case of economics, gender, and risk aversion. *Journal of Economic Methodology*, 21(3), 211-231.
- Nityananda, R. (2014). Fermi and the art of estimation. *Resonance*, 19(1), 73-81.
- O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law*, 15(4), 315.
- Page, S. (2007). Making the difference: Applying a logic of diversity. *Academy of Management Perspectives*, 21(4), 6-20.
- Page, S. (2008). The difference. In The Difference. Princeton University Press.
- Parmley, M. C. (2006). The effects of the confirmation bias on diagnostic decision making. Drexel University.
- Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541(7638), 532–535.
- Ray, R. (2006). Prediction markets and the financial" wisdom of crowds". *The Journal of Behavioral Finance*, 7(1), 2-4.

- Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences*, 13(10), 420–428.
- Röseler, L., Weber, L., Helgerth, K., Stich, E., Günther, M., Tegethoff, P., ... & Schütz, A. (2022). The Open Anchoring Quest Dataset: Anchored Estimates from 96 Studies on Anchoring Effects. *Journal of Open Psychology Data*, 10(1), 16.
- Sampedro, V., & Seoane Pérez, F. (2008). The 2008 Spanish general elections: “Antagonistic bipolarization” geared by presidential debates, partisanship, and media interests. The *International Journal of Press/Politics*, 13(3), 336-344.
- Schelling, T. C. (2006). Micromotives and macrobehavior. WW Norton & Company.
- Schneider, J. J. (2004). The influence of contrarians and opportunists on the stability of a democracy in the Sznajd model. *International Journal of Modern Physics C*, 15(05), 659-674.
- Schwämmele, V., González, M. C., Moreira, A. A., Andrade Jr, J. S., & Herrmann, H. J. (2007). Different topologies for a herding model of opinion. *Physical Review E*, 75(6), 066108.
- Shi, F., Tepeltskiy, M., Dueðe, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature human behaviour*, 3(4), 329-336.
- Shin, J. K., & Lorenz, J. (2010). Tipping diffusivity in information accumulation systems: More links, less consensus. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(06), P06005.
- Sznajd-Weron, K. (2005). Mean-field results for the two-component model. *Physical Review E*, 71(4), 046110.
- Starnini, M., Frasca, M., & Baronchelli, A. (2016). Emergence of metapopulations and echo chambers in mobile agents. *Scientific reports*, 6(1), 31834.

- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior research methods*, 42, 1096-1104.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.
- Surowiecki, J. (2005). The wisdom of crowds. Anchor.
- Sunstein, C. R. (1999). The law of group polarization. University of Chicago Law School, John M. Olin Law & Economics Working Paper, (91).
- Tessone, C. J., Toral, R., Amengual, P., Wio, H. S., & San Miguel, M. (2004). Neighborhood models of minority opinion spreading. *The European Physical Journal B-Condensed Matter and Complex Systems*, 39(4), 535-544.
- Travieso, G., & da Fontoura Costa, L. (2006). Spread of opinions and proportional voting. *Physical Review E*, 74(3), 036112.
- Tschann, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research*, 40(3), 271-300.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.
- Urbig, D., & Lorenz, J. (2007). Communication regimes in opinion dynamics: Changing the number of communicating agents. *arXiv preprint arXiv:0708.3334*.
- Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in cognitive sciences*, 22(3), 213-224.

Weidlich, W. (1991). Physics and social science—the approach of synergetics. *Physics reports*, 204(1), 1-163.

Weidlich, W. (2006). Sociodynamics: A systematic approach to mathematical modelling in the social sciences. Courier Corporation.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, 37(1), 62-69.

Zimmerman, F., Garbulsky, G., Ariely, D., Sigman, M., & Navajas, J. (2022). Political coherence and certainty as drivers of interpersonal liking over and above similarity. *Science Advances*, 8(6), eabk1909.

Zmigrod, L., & Goldenberg, A. (2021). Cognition and emotion in extreme political action: Individual differences and dynamic interactions. *Current Directions in Psychological Science*, 30(3), 218-227.

## 7. Apéndice 1

### Teorema de Diversidad Predictiva

El Teorema de Diversidad Predictiva (Page, 2007) permite expresar el error colectivo de una multitud ( $E$ , definido como la diferencia cuadrada entre la respuesta promedio de la multitud y la respuesta correcta) como dos componentes separadas: la diversidad predictiva ( $\delta$ ) definida como la diferencia cuadrada entre cada respuesta individual  $A$  y la respuesta promedio; y el error individual medio ( $\epsilon$ ), donde el error individual se define como la diferencia cuadrada entre las respuestas individuales de un participante y la respuesta correcta. La prueba de esta identidad matemática es directa:

Como se expresa, requerimos las siguientes definiciones:

$$\epsilon = \frac{1}{N} \sum_{q,i} (A^{(i)}_q - \theta_q)^2 = \frac{1}{N} \sum_{q,i} (A^{(i)}_q)^2 + \theta_q^2 - 2 \cdot \theta_q \cdot A^{(i)}_q, \quad [A1]$$

$$\delta = \frac{1}{N} \sum_{q,i} (A^{(i)}_q - \mu_q)^2 = \frac{1}{N} \sum_{q,i} (A^{(i)}_q)^2 + \mu_q^2 - 2 \cdot \mu_q \cdot A^{(i)}_q, \quad [A2]$$

$$E = \sum_q (\mu_q - \theta_q)^2 \quad [A3]$$

con  $N$  el tamaño de la población, el índice  $i$  recorriendo los  $N$  individuos, el índice  $q$  recorriendo las preguntas realizadas,  $\theta_q$  la respuesta correcta para cada pregunta,  $A^{(i)}_q$  la respuesta provista por cada individuo  $i$  en respuesta a la pregunta  $q$ , y  $\mu_q = \frac{1}{N} \sum_i A^{(i)}_q$  siendo la media de cada pregunta  $q$ .

Cuando restamos  $\epsilon$  y  $\delta$ , el término  $A^{(i)}_q$ <sup>2</sup> se cancela. Dado que los parámetros  $\mu_q$  y  $\theta_q$  son independientes de  $i$ , se obtiene:

$$\begin{aligned}\epsilon - \delta &= \frac{1}{N} \sum_{q,i} (\theta_q^2 - \mu_q^2 - 2 \cdot \theta_q \cdot A^{(i)}_q + 2 \cdot \mu_q \cdot A^{(i)}_q) = \\ &= \sum_q (\theta_q^2 - \mu_q^2) + \frac{2}{N} \sum_{q,i} (\mu_q \cdot A^{(i)}_q - \theta_q \cdot A^{(i)}_q) = \\ &= \sum_q (\theta_q^2 - \mu_q^2 + 2 \cdot \mu_q^2 - 2 \cdot \theta_q \cdot \mu_q) = \\ &= \sum_q (\mu_q - \theta_q)^2 = E\end{aligned}\quad [\text{A4}]$$

Esta identidad indica que el error colectivo  $E$  es exactamente igual al error individual medio menos la diversidad predictiva. Una implicancia de este resultado es que, en principio, el error colectivo puede reducirse no solo disminuyendo el error individual medio (en otras palabras, empleando o seleccionando individuos más sabios), sino también aumentando la diversidad predictiva de la multitud, que está relacionada con su varianza. Es importante tener en cuenta que esta identidad matemática se aplica a medidas de error convexas (relevantes para todos los problemas abordados en los capítulos 2 y 3), pero falla en el caso de medidas de error cónicas.

## Condiciones para Sobrepasar la Sabiduría de las Multitudes

Para encontrar el rango de parámetros bajo el cual el error colectivo de multitudes extremizadas es inferior al de la sabiduría de las multitudes clásica, se tomó la siguiente expresión como punto de partida:

$$|\mu_r - \theta| < |\mu - \theta| \quad [A5]$$

donde  $\mu_r$  es la media de las multitudes extremizadas,  $\mu$  es la media de la sabiduría de las multitudes, y  $\theta$  es la respuesta correcta. Suponiendo que  $\mu \neq \theta$  (es decir, la sabiduría de las multitudes no alcanza la respuesta correcta, lo que significa que hay margen para mejorar), se tiene:

$$\frac{|\mu_r - \theta|}{|\mu - \theta|} < 1 \quad [A6]$$

Adicionalmente, sin pérdida de generalidad, se asume que la multitud subestima la respuesta correcta, por lo cual  $\mu - \theta > 0$  (el procedimiento es análogo en el caso de sobreestimación).

Luego, se expande el valor absoluto  $\left| \frac{\mu_r - \theta}{\mu - \theta} \right|$ , y se multiplica por  $\mu - \theta$  de cada lado, llegando a:

$$-(\mu - \theta) < \mu_r - \theta < \mu - \theta \quad [A7]$$

Luego se resta  $\mu - \theta$  en los tres lados, y se llega a:

$$-2(\mu - \theta) < \mu_r - \mu < 0 \quad [A8]$$

Para mayor claridad, ahora se simplifica el problema y se asume un índice de anclaje constante  $w_0$ . Este caso particular es informativo, ya que un índice de anclaje constante produciría condiciones menos favorables para el modelo, dado que carecería del acceso al valor correcto  $\theta$  a través de la Ecuación [4] (la demostración para el caso general se puede encontrar a continuación). Si se tiene un índice de anclaje constante, utilizando las Ecuaciones [2] y [3], y reemplazándolas en la Ecuación [5], se obtiene:

$$\mu_r = w_0 \frac{A_L + A_H}{2} + (1 - w) \mu \quad [A9]$$

Si se define  $\bar{A} = \frac{A_L + A_H}{2}$  (valor medio de las anclas), y se reemplaza  $\mu_r$  en la Ec. [A8], se llega a:

$$-2(\mu - \theta) < w_0 \bar{A} + (1-w) \mu - \mu < 0 \quad [A10]$$

Ahora se divide por  $w_0$  y se suma  $\mu$  en todos lados, obteniendo:

$$-\frac{2}{w_0}(\mu - \theta) + \mu < \bar{A} < \mu \quad [A11]$$

La desigualdad en la Ec. [A11] implica que el rango de valores de  $\bar{A}$  para sobrepasar a la sabiduría de las multitudes tiene la forma:

$$\Delta_0 = \frac{2|\mu - \theta|}{w_0} \quad [A12]$$

Este resultado indica que las condiciones bajo las cuales el modelo mejora la sabiduría de las multitudes dependen del punto medio de las anclas (su promedio  $\bar{A}$ ). También implica que el rango de valores para  $\bar{A}$  que cumple esas condiciones es al menos el doble del tamaño del error.

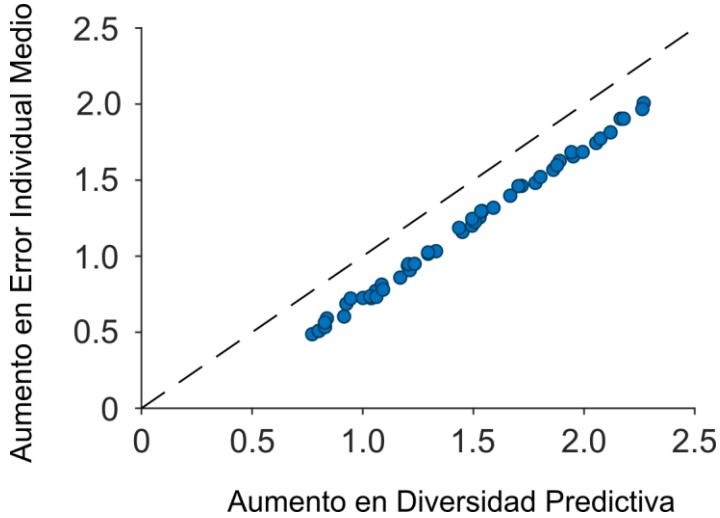
Si se elimina la suposición de un índice de anclaje constante, aún es posible seguir el mismo procedimiento, y llegar, en lugar de la Ecuación [A11], a la expresión general:

$$-\frac{4(\mu - \theta)}{(w_L + w_H)} + \mu + \frac{(w_L - w_H)}{(w_L + w_H)} \Delta A < \bar{A} < \mu + \frac{(w_L - w_H)}{(w_L + w_H)} \Delta A \quad [A13]$$

Esta expresión implica que el rango de valores donde las multitudes extremizadas superan a la sabiduría de las multitudes clásica está dado por la Ec. [6].

## **Competencia entre Aumentar Diversidad y Exactitud**

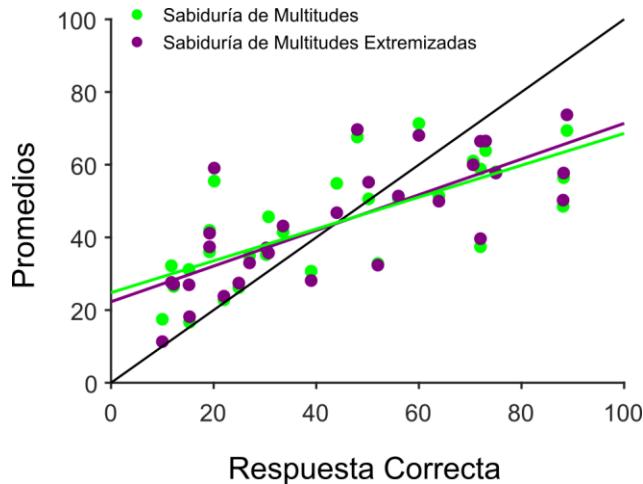
Dadas las predicciones del modelo y los resultados empíricos de los cuatro experimentos comportamentales, uno podría preguntarse si es posible aumentar la diversidad predictiva (a través del anclaje) mientras, al mismo tiempo, se reduce el error individual medio. En principio, no hay nada en el Teorema de Diversidad Predictiva que impida abordar ambos aspectos simultáneamente (es decir, se podría reducir el primer término  $\epsilon$  y aumentar el segundo término  $\delta$ ). Sin embargo, en todos los experimentos y simulaciones presentados aquí, se observó que cada vez que se reducía el error colectivo aumentando la diversidad, también se aumentaba el error individual medio. Para comprender mejor este fenómeno, se realizó un nuevo análisis basado en el modelo: se realizaron una serie de simulaciones del modelo propuesto, mientras se variaba sistemáticamente la distancia entre las anclas. Cada una de estas simulaciones produjo un conjunto de estimaciones con aproximadamente el mismo error colectivo, pero con diferente diversidad predictiva y error individual medio. A lo largo de las simulaciones, se encontró una asociación positiva entre esas dos variables (Fig. A1, coeficiente de correlación de Pearson,  $r=0.998$ ,  $p=2 \times 10^{-62}$ ), lo que sugiere que este enfoque establece un compromiso entre aumentar la exactitud y la diversidad, y que realmente compiten entre sí.



**Fig. A1. Competencia entre Aumentar Diversidad y Exactitud.** Simulaciones del modelo con cambios en el Error Individual Medio como función de cambios en la Diversidad Predictiva. Los parámetros del modelo son:  $w_0 = 0.2$  (peso cuando el ancla es la respuesta correcta),  $A = -1$  (promedio de las anclas),  $\mu = 0.2$  (respuesta promedio de la multitud),  $n = 10,000$  (tamaño de la multitud),  $\sigma = 1$  (desviación estándar de las respuestas de la multitud),  $\theta = 0$  (respuesta correcta). A lo largo de las simulaciones, se varió sistemáticamente la distancia entre las anclas desde  $\Delta = 5$  a  $\Delta = 10$  en pasos de 0.1. Los puntos azules muestran los resultados para diferentes especificaciones del modelo, y la línea punteada representa la identidad.

## Evaluación de Diferentes Métodos de Agregación

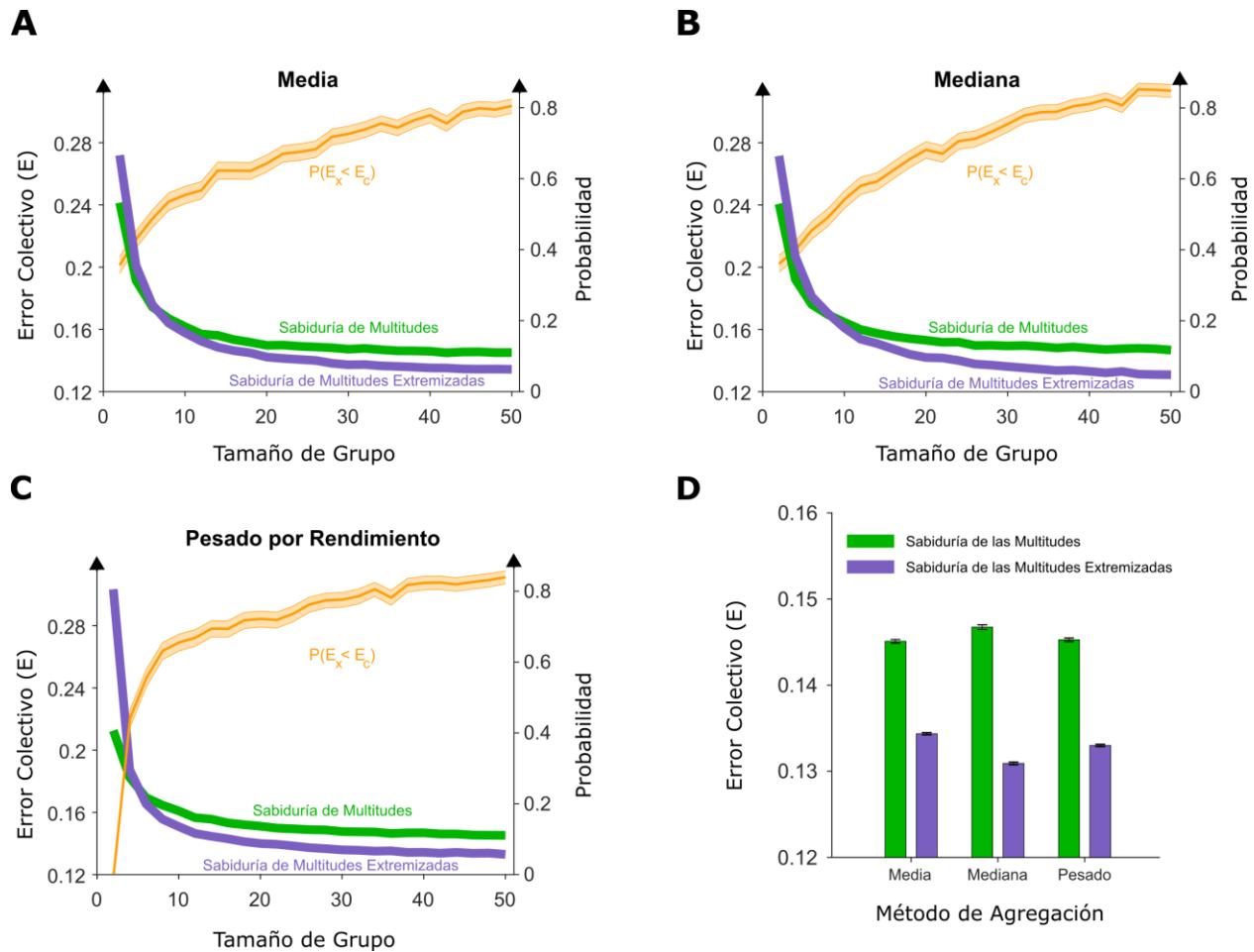
Al analizar los datos del Experimento 2, se encontró evidencia sólida consistente a favor de la existencia de un sesgo hacia probabilidades intermedias (Figura A2). Esto es cierto tanto para la sabiduría de las multitudes no extremizadas como para la sabiduría de las multitudes extremizadas.



**Fig. A2. Promedio de Respuestas vs. Respuesta Correcta.** Respuestas promedio para cada pregunta como función de la respuesta correcta, tanto para la sabiduría de multitudes no extremizadas (puntos verdes) como para la sabiduría de multitudes extremizadas (puntos violetas). Las líneas representan los mejores ajustes lineales.

Este resultado implica que, en principio, todavía hay margen para mejorar este método mediante la realización de diferentes procedimientos de agregación. Se probaron tres estrategias diferentes: las dos primeras son la media aritmética y la mediana, que han sido ampliamente utilizadas y comparadas desde la primera implementación de la sabiduría de las multitudes (Galton, 1907); la tercera es un promedio ponderado por el rendimiento (por ejemplo, ver Collins et al., 2023). En este último caso, se estimó el peso de cada participante utilizando un procedimiento de "dejar uno afuera" (leave-one-out): en primer lugar, se calculó el error individual medio de cada participante en función de las respuestas a todas las preguntas menos una; luego, se los normalizó para que sumen 1 al definir el peso de cada participante para la pregunta que quedó fuera.

Como se observa en la Figura A3, en todos los casos, este método resultó útil para reducir el error de la sabiduría de las multitudes no extremizadas (verde). Sin embargo, hay algunas diferencias entre los métodos de agregación. Usar la mediana en lugar de la media aumentó ligeramente el error de la sabiduría de las multitudes no extremizadas (específicamente, para el tamaño de multitud más grande N=50, prueba t no pareada:  $t(998)=2.37$ ,  $p=0.02$ ; tamaño del efecto:  $d$  de Cohen =  $0.16 \pm 0.09$ , 95% CI), pero redujo el error de las multitudes extremizadas (para el tamaño de multitud más grande N=50, prueba t no pareada:  $t(998)=6.38$ ,  $p=3\times10^{-10}$ ; tamaño del efecto:  $d$  de Cohen =  $0.35 \pm 0.09$ , 95% CI), lo que resultó en la mayor diferencia entre las dos multitudes. Los promedios ponderados por el rendimiento funcionaron mejor para multitudes más pequeñas, pero no mostraron diferencias estadísticas con la media para el tamaño de multitud más grande (prueba t no pareada:  $t(998)=0.33$ ,  $p=0.74$ ; tamaño del efecto:  $d$  de Cohen =  $0.02 \pm 0.09$ , 95% CI). En general, estos resultados respaldan aún más no sólo la robustez, sino también la versatilidad de este método, que puede utilizarse con diferentes estrategias de agregación.

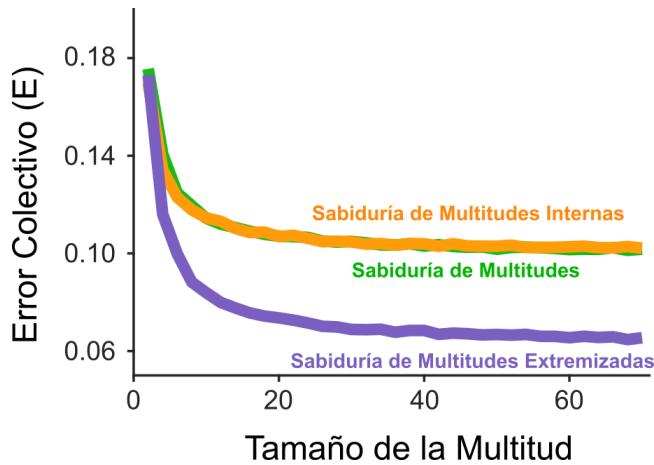


**Fig. A3. Comparación de los procedimientos de agregación para el Experimento 2.** (A-C) Error colectivo en función del tamaño de la multitud, para la sabiduría de las multitudes no extremizadas (verde) y las multitudes extremizadas (morado). El error estándar de las curvas se encuentra dentro del ancho de línea. La línea naranja corresponde a la probabilidad de que una muestra de la distribución de errores colectivos de la multitud extremizada (distribución violeta,  $E_x$ ) sea menor que una muestra de la distribución de errores colectivos de la multitud no extremizada (distribución verde,  $E_c$ ). (A) Método de agregación: media. (B) Método de agregación: mediana. (C) Método de agregación: media ponderada por el rendimiento. (D) Comparación del error colectivo de la multitud más grande ( $N=50$ ) para cada método de agregación. La sabiduría de las multitudes se presenta en verde, y la sabiduría de las multitudes extremizadas se presenta en violeta.

## Anclas Auto-Generadas y la Sabiduría de la Multitud Interna

Si bien el Experimento 4 demostró que el método propuesto puede ser efectivo –incluso sin conocimiento previo por parte de los experimentadores–, al pedir a los participantes que proporcionen sus propias anclas, existe un problema que sigue sin ser abordado: ¿podría el efecto de la sabiduría de la multitud interna (Herzog & Hertwig, 2009) explicar la eficacia del método? Si es así, se podría argumentar que no es el procedimiento de anclaje, sino este último efecto, el conductor principal de la reducción del error en la multitud. Para evaluar esto, y como se mencionó en la sección de Métodos del Experimento 4, se les solicitó a los participantes de la condición control que proporcionaran nuevas (diferentes) estimaciones para cada pregunta. Luego, se compararon la sabiduría de las multitudes, la sabiduría de las multitudes extremizadas, y la sabiduría de la multitud interna para cada tamaño de multitud posible (Figura A4).

Los resultados muestran que la sabiduría de la multitud interna (que, de acuerdo con Vul & Pashler (2008), mejora un 10% las estimaciones originales) no es estadísticamente distinta de la sabiduría de las multitudes en este experimento (para el tamaño de multitud más grande N=70, prueba t no pareada:  $t(998)=0.84$ ,  $p=0.40$ ; tamaño del efecto:  $d$  de Cohen =  $0.04 \pm 0.09$ , 95% CL; Factor de Bayes para la hipótesis nula: 18.01). Esto podría haberse esperado, ya que, al asumir una mejora promedio del 10% sobre las estimaciones individuales originales, para una multitud grande que involucra estimaciones de diferentes participantes, la mejora se vuelve insignificante en comparación con la varianza de las estimaciones en sí mismas. En general, estos resultados muestran que la eficacia de nuestro método no depende del efecto de la sabiduría de la multitud interna, sino que es en sí mismo un fenómeno diferente.



**Fig. A4. Comparación con la Sabiduría de la Multitud Interna para el Experimento 4.** Error colectivo en función del tamaño de la multitud, para la sabiduría de las multitudes no extremizadas (verde), la sabiduría de las multitudes extremizadas (violeta), y la sabiduría de la multitud interna (naranja). El error estándar de las curvas se encuentra dentro del ancho de las líneas.

## Definición del “Índice de Anclaje”

De las Ec. [2] y [3], se sigue que el índice de anclaje (Jacowitz & Kahneman, 1995) está definido como:

$$w_j = \frac{\mu_j - \mu}{A_j - \mu} \quad [16]$$

donde  $j$  representa  $L$  el ancla baja, o  $H$  para el ancla alta.

## Declaración de Prácticas Abiertas

Todos los datos y códigos que respaldan los análisis realizados en este capítulo están disponibles en el Open Science Framework en <https://osf.io/ch5qw/>.

El Experimento 2 fue pre-registrado en [https://aspredicted.org/RYC\\_4Y5](https://aspredicted.org/RYC_4Y5).

El Experimento 3 fue pre-registrado en [https://aspredicted.org/HZC\\_PTH](https://aspredicted.org/HZC_PTH).

El Experimento 4 fue pre-registrado en [https://aspredicted.org/YQ4\\_LYS](https://aspredicted.org/YQ4_LYS).

## Variables de Diseños Experimentales

Experimento	N	Número de Preguntas	Tipo de Preguntas	Rango de Respuestas	País	Muestra Representativa	Límite de Tiempo	Incentivo a la Exactitud	Pre-reg
1	120	14	Estimación	[0,+inf)	USA	No	Sí	Sí	No
2	396	30	Estimación	[0,100]	Argentina	No	Sí	No	Sí
3	620	2	Predicción	[0,+inf)	USA	Sí	No	Sí	Sí
4	226	10	Estimación	[0,+inf)	Argentina	No	Sí	No	Sí

**Tabla A1:** variables principales de diseño de los Experimentos 1-4.

## Experimento 1: preguntas y respuestas correctas

Question	Correct Answer
¿Cuál es la distancia (en millas) entre Memphis (Tennessee, EE. UU.) y Oklahoma City (Oklahoma, EE. UU.)?	421.51
¿Cuál es la distancia (en millas) entre Milwaukee (Wisconsin, EE. UU.) y Faith (Dakota del Sur, EE. UU.)?	714.18
¿Cuál es la distancia (en millas) entre Davenport (Iowa, EE. UU.) e Indianápolis (Indiana, EE. UU.)?	261.47
¿Cuál es la distancia (en millas) entre París (Francia) y Florencia (Italia)?	550.81
¿Cuál es la distancia (en millas) entre Atenas (Grecia) y Roma (Italia)?	652.97
¿Cuál es la altura (hasta la punta, en yardas) de la Torre Eiffel?	354.331
¿Cuál es la altura (hasta la punta, en yardas) del Monte Vesubio?	1400.92
¿Cuántos pisos tiene el Franklin Center?	60
¿Cuántas veces aparece la palabra "Jesús" en el Nuevo Testamento (Versión Internacional Nueva, sin distinguir mayúsculas y minúsculas)?	1273
¿Cuántas veces aparece la palabra "Alá" en el Corán (sin distinguir mayúsculas y minúsculas)?	2699
¿Cuántas veces aparece la palabra "Sabiduría" en el Corán (sin distinguir mayúsculas y minúsculas)?	50
¿Cuántos emperadores tuvo el Imperio Romano (sólo Imperio Unificado)?	71
¿Cuántos trasplantes de corazón se realizaron en 2016 en Estados Unidos?	3191
¿Cuántos puentes hay en París, Francia?	37

**Tabla A2:** preguntas evaluadas en el Experimento 1 y sus respuestas correctas.

## Experimento 2: preguntas y respuestas correctas

Question	Correct Answer
¿Qué porcentaje de la población argentina no está afiliado a ninguna religión?	12.2
¿Qué porcentaje de la población argentina de 20 años o más tiene sobrepeso u obesidad?	52.0
¿Qué porcentaje de miembros de la cámara de diputados de Argentina son mujeres?	39.0
¿Qué porcentaje de la población argentina tiene 14 años o menos?	24.9
Según una encuesta representativa hecha en 2013, ¿qué porcentaje de la población argentina cree que la homosexualidad no es moralmente aceptable?	27.0
Según una encuesta representativa hecha en 2013, ¿qué porcentaje de la población argentina cree que el aborto no es moralmente aceptable?	56.0
Según una encuesta representativa hecha en 2014, ¿qué porcentaje de la población argentina cree tener buena o muy buena salud?	75.0
¿Qué porcentaje de las muertes de varones de entre 15 y 24 años de edad, en Argentina, fueron por suicidio?	15.3
¿Qué porcentaje de la población argentina de 13 años o más tiene una cuenta de Facebook?	60.0
Según una encuesta representativa hecha en 2017, ¿qué porcentaje de la población argentina se cree capaz de distinguir noticias reales de "noticias falsas" (historias o hechos completamente inventados)?	72.0
Según el censo de 2010 en Argentina, ¿qué porcentaje de personas de entre 3 y 18 años de edad asiste a una institución educativa?	88.9
Según el censo de 2010 en Argentina, ¿qué porcentaje de la población habita en una vivienda sin descarga de agua o sin retrete?	15.2
Según el censo de 2010 en Argentina, contabilizando sólo mujeres mayores de 14 años, ¿qué porcentaje de ellas nunca tuvo un hijo?	30.3
Según el censo de 2010 en Argentina, contabilizando sólo personas mayores de 20 años, ¿qué porcentaje de la población tiene o está buscando trabajo?	70.6

Según el censo de 2010 en Argentina, ¿qué porcentaje de la población que habita en viviendas particulares tiene cobertura de salud?	63.9
¿Qué porcentaje de mujeres de entre 18 y 60 años de edad tienen empleo en Argentina?	50.2
¿Qué porcentaje de la población argentina tiene acceso a internet en su casa, ya sea a través de una computadora o dispositivo móvil?	73.0
Según una encuesta representativa hecha en 2013, ¿qué porcentaje de la población argentina cree que no es moralmente aceptable que dos adultos que no están casados tengan relaciones sexuales?	22.0
¿Qué porcentaje de la población de Argentina habita en una vivienda que le pertenece?	72.0
¿Qué porcentaje de la población argentina posee un teléfono inteligente (smartphone)?	48.0
Según una encuesta representativa del 2013, ¿qué porcentaje de la población argentina cree que la mayor parte de la gente es confiable?	19.2
¿Qué porcentaje de la población argentina cree que algunas vacunas pueden causar autismo en niños sanos?	10.0
En Argentina, ¿qué porcentaje del total de riqueza inmobiliaria pertenece al 1% más rico de la población?	44.0
Según el censo de 2010 en Argentina, ¿qué porcentaje de personas mayores de 20 años completó la educación secundaria?	20.1
Según el censo de 2010 en Argentina, ¿qué porcentaje de la población mayor de 80 años nació en otro país?	11.8
Según una encuesta representativa del 2013, ¿qué porcentaje de la población argentina no está para nada interesado en política?	30.7
Según el censo de 2010 en Argentina, ¿qué porcentaje de la población que vive en la provincia de Santa Fe es menor de 65 años de edad?	88.2
Según el censo de 2010 en Argentina, ¿qué porcentaje de la población es soltera en su estado civil, y no está en pareja?	33.5
Según el censo de 2010 en Argentina, ¿qué porcentaje de la población no habita en un departamento?	88.1
Según el censo de 2010 en Argentina, ¿qué porcentaje de las personas de entre 20 y 29 años de edad trabajan y estudian?	19.2

**Tabla A3:** preguntas evaluadas en el Experimento 2 y sus respuestas correctas.

### **Experimento 3: preguntas y respuestas correctas**

Question	Correct Answer
¿Cuántas nuevas muertes por COVID-19 crees que habrá en los Estados Unidos en la próxima semana (27 de julio - 2 de agosto)?	7987
¿Cuántos nuevos casos de COVID-19 crees que habrá en los Estados Unidos en la próxima semana (27 de julio - 2 de agosto)?	442,417

**Tabla A4:** preguntas evaluadas en el Experimento 3 y sus respuestas correctas.

### **Experimento 4: preguntas y respuestas correctas**

Pregunta	Respuesta Correcta
¿Cuál es la altura del “Obelisco de Buenos Aires”?	67
¿Cuál es la altura de la “Catedral de La Plata”?	112
¿Cuál es la altura del “Palacio de Aguas Corrientes”?	21
¿Cuál es la altura del “Palacio del Congreso de la Nación Argentina”?	80
¿Cuál es la altura de la “Torre de los Ingleses de Retiro”?	60
¿Cuál es la altura de la Torre Eiffel (a la punta)?	330
¿Cuál es la altura del Empire State Building (a la punta)?	443
¿Cuál es la altura del Buckingham Palace?	24
¿Cuál es la altura de la Torre de Shanghai?	632
¿Cuál es la altura de la Torre de Pisa?	56

**Tabla A5:** preguntas evaluadas en el Experimento 4 y sus respuestas correctas

## 8. Apéndice 2

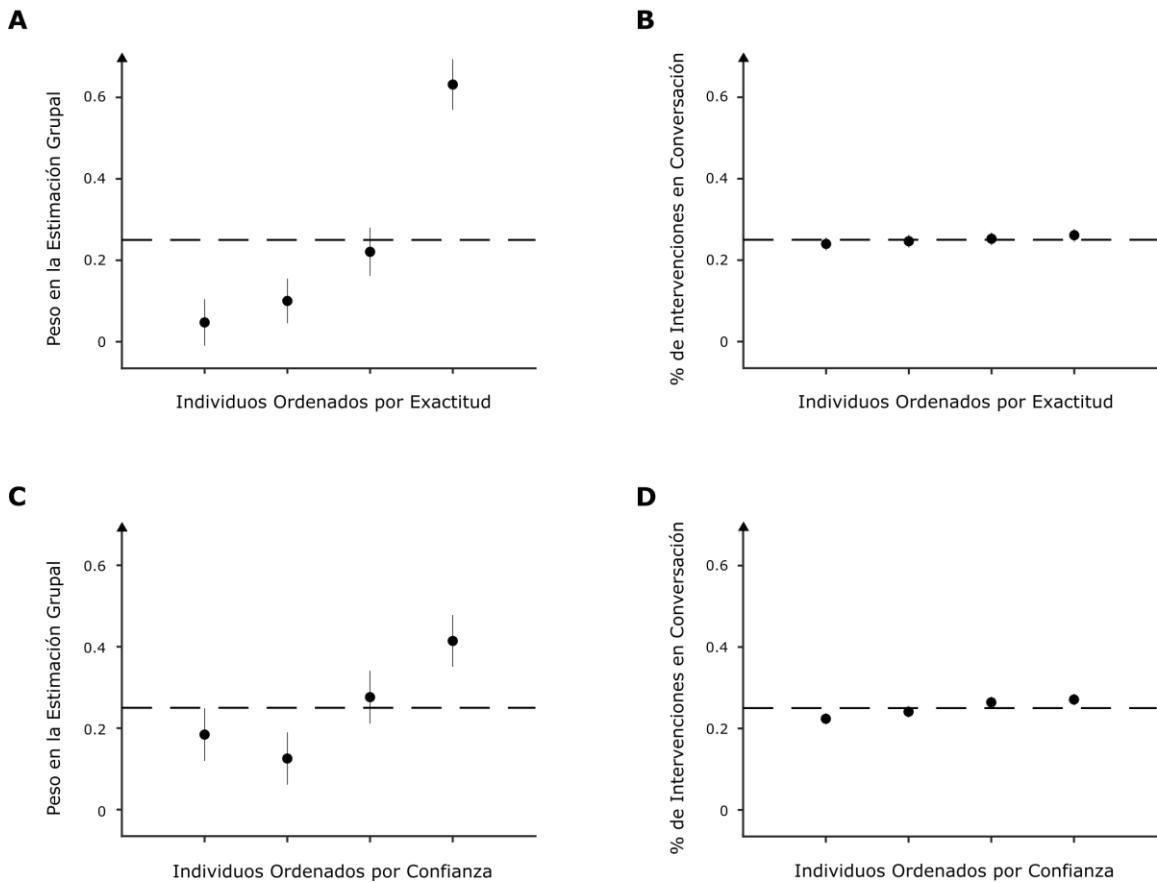
Preguntas realizadas en los distintos experimentos del Capítulo 3.

Preguntas	Respuesta Correcta
¿Cuántos goles se metieron en total en el Mundial 2014 de la FIFA jugado en Brasil?	171
¿Cuántos escalones tiene la escalera interna del Obelisco de Buenos Aires?	206
¿Cuántas personas (en millones) viven en América del Sur?	373
¿Cuál es la longitud (en metros) del recorrido total de la Línea A del subte de Buenos Aires?	10800
Hasta el 2021, en cuántas finales de la Copa Libertadores ha participado un equipo argentino?	37
¿Cuántos kilos de helado consume una persona por año en Argentina?	7
¿Cuántos capítulos tiene la serie de televisión "Friends"?	236
¿Cuántos países tienen territorio en el hemisferio sur, sin contar al continente Antártico?	48

**Tabla S1:** preguntas evaluadas en los tres experimentos y sus respuestas correctas.

## Figura Suplementaria 1

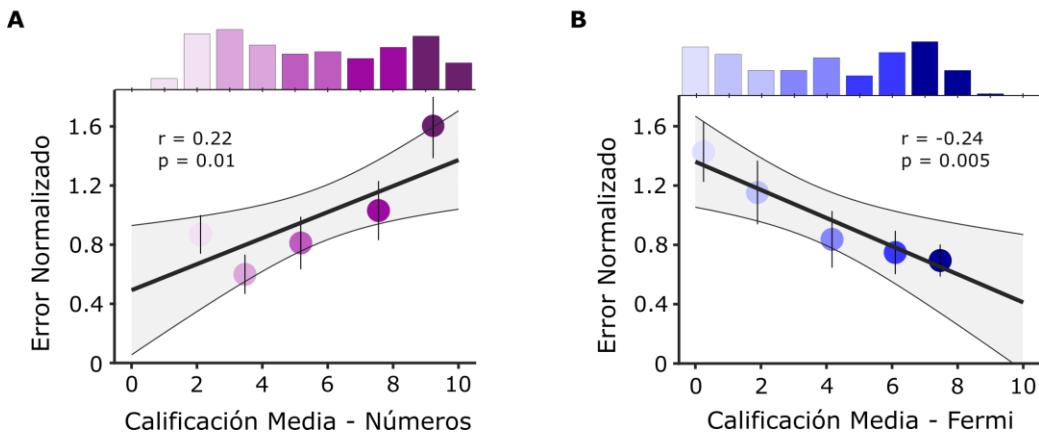
### Experimento 1 (N=520)



**Fig. A5. Impacto individual de cada individuo sobre las estimaciones grupales y el porcentaje de intervenciones para el Experimento 1.** (A) Peso en la estimación grupal de cada individuo ordenados por exactitud. (B) Porcentaje de intervenciones (mensajes) de cada individuo ordenados por exactitud. (C) Peso en la estimación grupal de cada individuo ordenados por confianza. (D) Porcentaje de intervenciones (mensajes) de cada individuo ordenados por confianza.

## Figura Suplementaria 2

### Experimento 2 (N=240)



**Fig. A6. Análisis de Conversaciones del Experimento 2.** (A) Error normalizado en función de las calificaciones de la estrategia de “Números” (promediadas entre evaluadores). Se representaron los cuartiles de la distribución de errores normalizados, junto con el mejor ajuste lineal de los datos, y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También se exhibe el coeficiente de correlación de Pearson y el p-valor correspondiente. (B) Error normalizado en función de las calificaciones de la estrategia “Fermi” (promediadas entre evaluadores). Se representan los cuartiles de la distribución de errores normalizados, junto con el mejor ajuste lineal de los datos y los intervalos de confianza del 95%. Las barras sobre el gráfico principal muestran la distribución real de errores normalizados. También se exhibe el coeficiente de correlación de Pearson y el p-valor correspondiente.

## **9. Apéndice 3**

### **Disponibilidad de Datos del Capítulo 4**

Los datos que respaldan los descubrimientos presentados en este studio son de acceso abierto, disponibles en F. Barrera Lemarchand (2020). “*Polarizing crowds: Consensus and bipolarization in a persuasive arguments model*”, GitHub. <https://github.com/fedex192/Pol-crowds-c-b-in-PAM>.”