

EL SESGO DE CONFIRMACIÓN COMO INCONMENSURABILIDAD.
MICROFUNDAMENTANDO EL “EFECTO CONTRARIAN”

Matías Daniel Fernández¹ – (Universidad Nacional de La Plata)

daniel.mza@gmail.com

Abstract

La posibilidad de trabajar con modelos cada vez más complejos conllevó a la detección de patrones emergentes difíciles de explicar y hasta de interpretar, demostrando que los modelos sencillos aún tienen un potencial explicativo importante.

Uno de los patrones emergentes de mayor impacto en las ciencias sociales, ha sido, quizás, la revelación que la polarización parece tener un mayor protagonismo en el escenario mundial que el consenso. El efecto *contrarian*, como fenómeno macroscópico, no es más que la presencia de agentes que van “contra la corriente”, es decir, que basan sus propias valoraciones personales en las antípodas de las construcciones hegemónicas de la sociedad, posibilitando explicaciones a estos fenómenos generalizados.

El presente trabajo busca acercarse a otra interpretación posible sobre el comportamiento microscópico del efecto *contrarian*, a partir de la *inconmensurabilidad* de los paradigmas como manifestación concreta del *sesgo de confirmación*. Por lo tanto, ya no existe una proporción p de agentes que se comportan como *contrarians*, sino que ahora, existe una probabilidad p de que cualquier agente pueda mostrar dificultades en los mecanismos de lenguaje y se comporte en ese caso particular como un *contrarian*.

Tanto si consideramos esta probabilidad endógena como exógena, vemos que el sistema puede generar problemas de lenguaje que hace imposible a los agentes llegar a escenarios de consenso, o hacerlo con mayores dificultades. De hecho, las posiciones neutrales resultan, finalmente, claves para alcanzar, en algunos casos, escenarios de consenso.

JEL: D7, D83

¹ Licenciado en Economía (Universidad Nacional de Cuyo), Licenciado en Ciencia Política y Administración Pública (Universidad Nacional de Cuyo), estudiante de la Maestría en Economía (Universidad Nacional de La Plata).

Introducción

El Efecto Contrarian

Hoy, la ciencia está cambiando a pasos acelerados. El acceso a grandes cantidades de datos y la capacidad de procesarlos en cuestión de minutos, ha permitido a los analistas trabajar con modelos cada vez más complejos. No obstante, esta particularidad también conllevó al surgimiento de patrones emergentes difíciles de explicar y hasta de interpretar, demostrando que los modelos sencillos aún tienen un potencial explicativo importante (SAN MIGUEL, 2016).

Uno de los patrones emergentes de mayor impacto en las ciencias sociales, al menos en este último tiempo, ha sido, quizás, la revelación que la polarización parece tener un mayor protagonismo en el escenario mundial que el consenso (FLACHE, 2016). Desde elecciones presidenciales (Estados Unidos, España, Austria, Argentina, Francia) y el Brexit hasta la formación de opinión en la sociedad, los ejemplos se repiten y han cobrado fuerza durante estos últimos diez años. Sin embargo, gran parte de la literatura a tal efecto está repleta de modelaciones y preposiciones que indican que la interacción humana, aun la que se desarrolla a la sombra de los medios masivos de comunicación, tienden a generar procesos de consenso y *retorno a la media*.

Sin embargo, a nivel macroscópico, el **efecto contrarian** ha sido una de las adiciones más importante en los modelos tradicionales y un elemento que ha sido capaz de mostrar cómo el consenso y la polarización son estados de un mismo sistema, que cambia a medida que la interacción de los agentes se va desarrollando a lo largo del tiempo (GALAM, 2004). Actualmente es un tema muy estudiado, y distintas aplicaciones van echando luz sobre un problema reciente, pero que ha comenzado a preocupar a muchos analistas sociales que ven como estos fenómenos de polarización han comenzado a fragmentar el tejido social en muchos lugares del mundo.

Por su parte, el efecto *contrarian*, como fenómeno macroscópico, no es más que la presencia de agentes que van “contra la corriente”, es decir, que basan sus propias valoraciones personales en las antípodas de las construcciones hegemónicas de la sociedad. De esta manera, cuando la mayoría de los agentes de una comunidad se posicionan a favor de A, estos agentes comenzarán a dar el debate hacia el interior de la sociedad en apoyo a B, pero si esta última opción se

vuelve mayoritariamente en la favorita, comenzará a dedicar sus esfuerzos en la defensa de A.

Si bien este elemento macroscópico ha sido una incorporación por demás interesante para la reinterpretación de muchos fenómenos, como burbujas de precios, volatilidad en los mercados, procesos electorales cerrados, votaciones inesperadas, entre otras; la realidad es que aún se está trabajando en la conceptualización microscópica del efecto *contrarian*:

- ϕ ¿Cómo es que existen agentes que “padecen” de dicho fenómeno mientras que otros no?
- ϕ ¿Cómo operan sus valoraciones y su comportamiento como agente?
- ϕ ¿Cuáles son sus intereses en la interacción social?

Éstas y muchas otras preguntas están siendo estudiadas por diversas disciplinas, porque también puede constituirse en una nueva punta de investigación que permita construir un nuevo marco teórico sobre el comportamiento de grandes masas.

Sesgo de confirmación

El presente trabajo busca acercarse a otra interpretación posible sobre el comportamiento microscópico del efecto *contrarian*. En este sentido, no diremos que una proporción p de los agentes padecen del fenómeno, sino que todos y cada uno de los agentes son plausibles de estar actuando como un *contrarian* con una probabilidad p_i , lo que nos llevará a la necesaria pregunta sobre qué variables están involucradas en ello.

A tal efecto, la propuesta se basa en el concepto de **sesgo de confirmación** (RABIN & SHRAG, 1999), por el cual, cuando un agente escucha un argumento contrario a su opinión actual, existe una probabilidad p_i que pueda malinterpretar la información que recibe, que ésta le produzca algún tipo de rechazo, o que opere en su mente de alguna forma que en lugar de acercarlo a una posición de consenso sólo sirva para confirmar su propia concepción inicial, polarizando aún más el sistema en su conjunto. Aún más, puede considerarse que la probabilidad p_i depende tanto de la intensidad argumentativa del propio agente como la de su interlocutor, lo que genera asimetrías dentro del sistema entendido desde una perspectiva holística.

Paradigmas y revoluciones científicas

El concepto de paradigma dentro de la ciencia sirve para explicar cómo se generan consensos en torno a visiones filosóficas, epistemológicas, metodológicas o interpretativas, que sirven para construir marcos teóricos y supuestos iniciales sobre los cuales el resto de la ciencia se construye, muchas veces, de manera lineal (KUHN, 1962).

Esta visión ha servido para explicar cómo las distintas disciplinas, sobre todo las exactas y físico-naturales, han sabido producir a lo largo de la historia visiones unificadas sobre cómo interpretar la realidad, que no necesariamente muestran correlación y coherencia entre sí. De esta manera, ya nadie recuerda la teoría del *flogisto* en la química general, y concuerda que el paradigma actual puede explicar mejor los fenómenos de la misma (KUHN, 1962).

Esta visión sobre la construcción científica incorpora otro concepto importante, el de **incommensurabilidad**. Esta categoría indica que dos paradigmas se construyen a partir de cosmovisiones distintas a partir de lenguajes y categorías diferenciadas, por lo que sería normal que las distintas teorías no pudieran interpretarse la una a la otra. Este fenómeno encuentra correlato con el sesgo de confirmación, en tanto un paradigma, frente a los argumentos del otro, carece de herramientas para interpretarlo, así como también detectar anomalías o construir soluciones sistemáticas, por lo que recurre normalmente a respuestas *ad-hoc*.

Thomas Khun (1962) propone que las revoluciones científicas se producen a medida que sectores de la comunidad científica construyen una nueva visión sobre el mundo, que intentan poner en el centro de la escena a través de la persuasión y mecanismos de consenso. El autor advierte una larga serie de dificultades para que se produzca el intercambio entre el paradigma que no termina de morir y aquel que no termina de nacer, e indica que no dependerá tanto de la capacidad explicativa del nuevo marco teórico, sino de la capacidad que tienen sus promotores para demostrar los potenciales de adoptarlo frente a las posibles desventajas. De cualquier manera, la historia ha demostrado que mejores marcos explicativos han tenido la capacidad de instalarse rápidamente en los sistemas científico-tecnológicos, en tanto redujeron la presencia de anomalías y favorecieron a grupos de investigación para empujar la frontera de conocimiento. En estos términos, resulta clara la analogía entre la noción de revoluciones científicas del autor y los

mecanismos de adopción de innovaciones de las sociedades contemporáneas (IGLESIAS, 2016).

Sin embargo, en una larga lista de disciplinas, muchas de ellas asociadas a las ciencias sociales y humanas, aún se discute si rige una visión paradigmática de la ciencia. Por lo tanto, se impulsan por separado dos caracterizaciones distintas sobre la comunidad científica. Por un lado, una visión aparadigmática, que se sustenta en la creencia de un carácter pre-científico de estas disciplinas, en tanto no han consolidado una visión filosófica, epistemológica, metodológica y teórica como han logrado otras disciplinas; y, por otro lado, una visión multi-paradigmática, que explica la existencia de disciplinas donde no existe un consenso, básicamente, porque prima la propia polarización de las visiones científicas, las cuales conviven en una constante disputa.

En el presente trabajo, se buscará utilizar esta visión multi-paradigmática, como ejemplo, para tratar de explicar cómo la inconmensurabilidad y el sesgo de confirmación son capaces de crear un estado de polarización latente entre dos visiones distintas, y encontrar si es posible que en algún momento pueda generarse algún tipo de visión compartida.

El modelo

Descripción general

El modelo consta de N agentes, que interactúan en grupos de a dos personas durante T períodos consecutivos. Cada uno de los agentes puede adscribir al paradigma A o B con una intensidad que denominaremos a_i y b_i , respectivamente, de forma tal que $a_i = -b_i$ (es decir, los paradigmas son simétricamente opuestos). Los agentes envían las señales $\alpha_i = a_i$, cuando adscriben al paradigma A , y $\beta_i = b_i$, cuando adscriben al paradigma B . No hay problemas en la emisión del mensaje. De esta manera, puede darse dos tipos de interacciones entre los agentes i y j :

- ♠ Cuando $a_i, a_j \geq 0$ o $a_i, a_j \leq 0$, ambos comparten el mismo paradigma (o poseen una versión neutra sobre los mismos) y no tienen problemas a la hora de interpretar el mensaje del otro, por lo que se genera una solución de compromiso:

$$a_i = a_j = \frac{\alpha_i + \alpha_j}{2}$$

ϕ En cambio, cuando a_i y a_j son de signo opuesto, entra en juego el principio de inconmensurabilidad descrito anteriormente. En este sentido, los agentes pueden no entender las categorías utilizadas por el otro, lo que da lugar a posibles respuestas que van en detrimento del consenso. En este caso, modelamos ese factor a partir de valores p_i y p_j , que indican la probabilidad que cualquiera de los agentes interprete el mensaje como si fuera de signo opuesto, es decir, el agente comprende β_j cuando el mensaje original había sido a_j . Este mecanismo engloba distintos fenómenos que pueden estar sucediendo en la mente de los agentes, por ejemplo, los agentes pueden estar malinterpretando las señales enviadas por el otro; entender correctamente el mensaje, pero generar un rechazo hacia el contenido de la argumentación, lo que los lleva a confirmar su posición inicial; no comprender el contenido del mensaje y sustituirlo por su propia valoración, entre otras.

Tras esta situación, ambos agentes utilizarán la solución de compromiso propuesta anteriormente, salvando la posibilidad que opere el cambio de signo descrito. Por su parte, este cambio de signo operará con la siguiente probabilidad:

$$p_i = \begin{cases} (\alpha_i - \alpha_j)^2, & \alpha_i - \alpha_j < 1 \\ 1, & \alpha_i - \alpha_j \geq 1 \end{cases}, \text{ cuando } p_i \text{ se determina endógenamente y,}$$

$$p_i = q, \text{ cuando } p_i \text{ se determina exógenamente}$$

Las fórmulas planteadas buscan transmitir el propio concepto de inconmensurabilidad. En este sentido, una mayor intensidad a_i , indica una mayor especialización dentro de un paradigma, y por lo tanto una adecuación a las categorías que este presenta. Es decir, un agente que adhiere a un paradigma, está entrenado en la cosmovisión correspondiente y, por lo tanto, es más proclive a utilizar categorías específicas del mismo en su lenguaje y argumentos. Entonces, como el agente no está preparado para comprender categorías distintas a las que ya está familiarizado, tendrá dificultades tanto para comunicar adecuadamente como para comprender los argumentos de agentes que adscriben a un paradigma distinto. En resumen, una mayor intensidad dentro de un paradigma significa, para el agente, mayores dificultades en comprender a miembros del paradigma opuesto, así como también mayor problemas de los demás agentes en comprenderlo a él.

Metodología

En el presente trabajo se busca conocer la interacción de los agentes a través del análisis estadístico de simulaciones realizadas con *Jupyter Notebook* (lenguaje *Python*), utilizando las librerías *numpy*, *simpy*, *stats* y *pandas*.

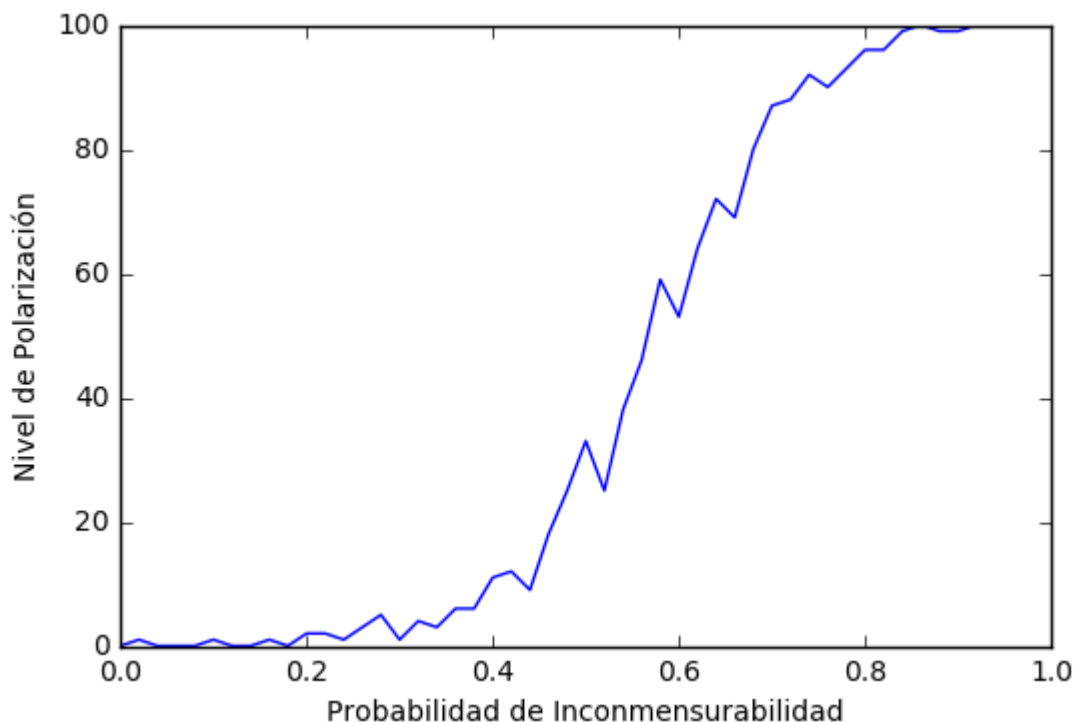
Se realizarán R simulaciones diferentes con T períodos cada una, donde la posición a_i del agente i se determinará aleatoriamente mediante una distribución uniforme. Allí, los N agentes, en cada período, serán emparejados aleatoriamente. Cuando a_i y a_j tienen signo opuesto, existe una probabilidad p_i y p_j que se produzca la inconmensurabilidad, lo que permite computar el resultado aleatoriamente mediante una distribución uniforme.

En función de los α_i y α_j que hayan sido interpretados por cada uno de los agentes, se producirán soluciones de compromiso como las descritas anteriormente, posibilitando tanto equilibrios de consenso como de polarización.

Simulaciones: p_i exógena

En este apartado se simulará **un modelo donde la probabilidad p_i será exógena** e igual a q , lo que permitirá evaluar la sensibilidad de la proporción ρ de equilibrios de polarización en función del valor de q , y determinar el umbral a partir del cual $\rho(q) > 0$ de manera significativa.

Resultados



Como era de esperarse, existe una correlación positiva entre q y ρ , lo que demuestra que la polarización es producto del sesgo de confirmación que se suscita entre los agentes. Como aquí la probabilidad de inconmensurabilidad es exógena, a medida que ésta aumenta, el sesgo de confirmación se acrecienta, y una misma situación inicial alcanza equilibrios completamente diferentes.

El umbral a partir del cual $\rho(q) > 0$ oscila para valores de $q \approx 0.4$, mientras que para valores de $q > 0.88$, tenemos que $\rho(q) \approx 1$.

Simulaciones: p_i endógena

En este apartado se simulará **un modelo donde la probabilidad p_i será computada endógenamente**, es decir, habrá una proporción p de las R simulaciones que resulten en equilibrios de polarización, mientras que una proporción $1 - p$ resultará en equilibrios de consenso. Además, podremos determinar la probabilidad promedio y la probabilidad de inconmensurabilidad:

ϕ Probabilidad de inconmensurabilidad

$P(\alpha_j = -\alpha_i | \text{sgn}(\alpha_i) \neq \text{sgn}(\alpha_j))$: Este valor es la probabilidad de observar un mensaje α_i , emitido por el agente i , que fue interpretado como $-\alpha_i$ por el agente j , partiendo de la situación que α_i y α_j poseen signo opuesto. Se calcula como la media únicamente de las probabilidades positivas.

ϕ **probabilidad promedio** $P(\alpha_j = -\alpha_i)$: siempre que α_i y α_j posean el mismo signo, la probabilidad de inconmensurabilidad es 0, en tanto los agentes poseen un marco teórico común que abona el entendimiento y el consenso. No obstante, cuando las mismas posean signo opuesto, se calculan endógenamente las probabilidades p_i y p_j , tal y como hemos descrito anteriormente. Este valor es la probabilidad de observar un mensaje α_i , emitido por el agente i , que fue interpretado como $-\alpha_i$ por el agente j . Se calcula como la media de todas las probabilidades, incluyendo cuando $p_i = 0$ y $p_j = 0$.

Para ello se realizarán 20 simulaciones diferentes con ánimos de contrastar los resultados encontrados.

Resultados

Caso	% Polarización	p promedio	p incomensurabilidad
15	0.97	0.530879	0.986670
18	1.00	0.528024	0.999822
06	0.93	0.510671	0.971515
01	0.97	0.507725	0.988562
05	0.74	0.422400	0.900352
10	0.89	0.419120	0.959940
17	0.45	0.251787	0.815087
19	0.86	0.239617	0.950553
20	0.05	0.107331	0.567035
16	0.05	0.098905	0.543480
09	0.05	0.095261	0.705545
07	0.02	0.089590	0.539184
08	0.00	0.057420	0.501703
14	0.00	0.049966	0.427173
04	0.00	0.042636	0.442648
12	0.00	0.039720	0.428146
11	0.00	0.036817	0.568005
02	0.00	0.034978	0.358332
03	0.00	0.026180	0.295051
13	0.00	0.009957	0.302991

Como puede apreciarse en la tabla anterior, cuando p_i es endógena hay dos indicadores que muestran la probabilidad para que una situación inicial se polarice, que impactan en distinta magnitud. Por un lado, la probabilidad promedio oscila entre valores cercanos a 0.00 y 0.60, con una incidencia significativa a partir de valores superiores a 0.20².

Esto demuestra que aquellas situaciones que inicialmente son más polarizadas, y por lo tanto hay menos probabilidad que a_i y a_j compartan el signo, tienden a presentar situaciones finales donde esta propiedad se acrecienta, mientras que situaciones iniciales menos polarizadas tienden con mayor probabilidad a

² En los resultados se muestra una tabla donde se ordenan 20 casos simulados. Si se repite varias veces el algoritmo, estos resultados cambian, pero reafirman y clarifican los valores sobre lo que se hace referencia.

situaciones de consenso, existiendo incluso casos donde nunca se llega a polarizar. Por otra parte, la probabilidad de inconmensurabilidad, que es la que determina la probabilidad de sesgo de confirmación en aquellos casos donde a_i y a_j presentan signos opuestos, impacta también positivamente en la probabilidad ρ de alcanzar escenarios finales polarizados.

Esto es producto de situaciones iniciales con posiciones más extremas, que acrecienta la determinación endógena de p_i . En cambio, en situaciones iniciales donde los valores tienden a concentrarse alrededor de $a_i = 0$, vemos que aún existe posibilidad de consenso, dado que la probabilidad de inconmensurabilidad es más baja. En este sentido es importante recordar que $p_i[(a_i - a_j) > 1] = 1$, lo que produce que posiciones extremas de algunos agentes tiendan a aumentar la probabilidad de inconmensurabilidad de todo el sistema.

Conclusiones

La **inconmensurabilidad**, como una manifestación posible del **sesgo de confirmación**, aparece como un fundamento microscópico del **efecto contrarian** que se discutió anteriormente. No obstante, la conclusión macroscópica puede abordarse desde una perspectiva distinta, ya que no hablamos más de una proporción p de agentes que se comportan como *contrarians*, sino que ahora, existe una probabilidad p de que cualquier agente pueda mostrar dificultades en los mecanismos de lenguaje y se comporte en ese caso particular como un *contrarian*. Esta última conclusión es importante, pues ya no necesitamos elementos que nos permitan justificar la existencia de agentes de este tipo, como resulta en los casos de adopción de innovaciones (IGLESIAS, 2016).

Cuando se considera que p_i es exógena, vemos que ρ y q se encuentran correlacionados positivamente, lo que acompaña la intuición que mayores problemas de lenguaje impactan negativamente en los escenarios de consenso. No obstante, cuando p_i es endógena, vemos que no sólo p y ρ se encuentran correlacionados positivamente, sino que se advierte que p puede ser entendida como una probabilidad promedio de problemas de interpretación y comprensión en general, sino también como una probabilidad de inconmensurabilidad, donde esta última aparece como más importante y con un impacto más directo sobre los resultados. Esto conlleva otra conclusión fuertemente intuitiva, en tanto la polarización será más frecuente a medida que las posiciones de los paradigmas sean más extremas, pero

las posiciones neutrales resultan claves para alcanzar, en algunos casos, escenarios de consenso.

Bibliografía

- FLACHE, A. (2016). The complexity of social integration in a diverse society. Can we connect theoretical models and big data? *SocioFísica Argentina* (págs. 1-35). Capital Federal: FCEyN - UBA.
- GALAM, S. (2004). Contrarian Deterministic Effect: the "Hung Elections Scenario". *Physica A: Statistical Mechanics and its Applications*, 453-460.
- KAHNEMAN, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 1449-1475.
- KAHNEMAN, D., & TVERSKY, A. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 1124-1131.
- KUHN, T. (1962). *La estructura de las revoluciones científicas*. Chicago: University of Chicago Press.
- RABIN, M., & SHRAG, J. (1999). First Impressions Matter: A Model of Confirmatory Bias. *Quarterly Journal of Economics*, 37-82.
- SAN MIGUEL, M. (2016). What can we learn from simple models of social interaction? *SocioFísica Argentina* (pp. 1-43). Capital Federal: FCEyN - UBA.