

EL UNIVERSO Y SUS OBSERVADORES

IDEAS FUNDAMENTALES

Edición científica

IDE-001 a IDE-024

Juan Segurola

Versión 0.3 · 19 de septiembre de 2026

© 2026 Juan Segurola. Todos los derechos reservados.

Sobre esta edición

Este volumen compila los 24 artículos en versión de publicación de la serie Ideas fundamentales. Cada artículo se reproduce a partir del corpus IDE v0.83 contrafirmado. La compilación no altera las afirmaciones, las referencias, las etiquetas epistémicas ni las versiones de los artículos.

Índice

Artículo	Título	Versión de origen	Página
IDE-001	Energía: una cuenta conservada, no una sustancia	v0.8	4
IDE-002	Información: cuantificar alternativas y dependencia	v0.6	11
IDE-003	Entropía: conteo, restricciones e irreversibilidad	v0.8	17
IDE-004	Causalidad: intervención, contrafactuales e influencia física	v0.7	23
IDE-005	Simetría: transformaciones, invariantes y estados rotos	v0.5	29
IDE-006	Escala: ¿qué detalles importan en cada nivel?	v0.7	35
IDE-007	Emergencia: cuando los patrones colectivos se vuelven explicativos	v0.8	41
IDE-008	Probabilidad: un cálculo para alternativas inciertas	v0.5	47
IDE-009	Campo: magnitudes distribuidas en un dominio	v0.5	53
IDE-010	Retroalimentación: cuando la salida vuelve como entrada	v0.6	58
IDE-011	Selección: cómo la persistencia diferencial acumula estructura	v0.7	63
IDE-012	Representación: cuando una cosa está por otra	v0.9	68
IDE-013	Tiempo: relojes, orden y la flecha	v0.10	73
IDE-014	Medición: de la indicación al	v0.6	78

Artículo	Título	Versión de origen	Página
	valor		
IDE-015	Modelo: una máquina de inferencia con dominio	v0.8	83
IDE-016	Complejidad: varias preguntas bajo una palabra	v0.5	88
IDE-017	Incertidumbre: nombrar las fuentes antes de sumarlas	v0.5	93
IDE-018	Conexión: qué afirma una arista	v0.6	98
IDE-019	¿Qué es la luz? Onda, cuanto y observador	v0.9	103
IDE-020	¿Qué es una partícula?	v0.12	108
IDE-021	¿Qué es un campo?	v0.8	113
IDE-022	¿Qué significa la masa?	v0.7	118
IDE-023	¿Qué es la observación?	v0.20	123
IDE-024	¿Qué es una señal?	v0.9	128

Energía: una cuenta conservada, no una sustancia

IDE-001 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.8 ES · 17 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Multiescala, de la mecánica y la química a la cosmología

Límite: La energía como magnitud de balance definida y conservada; la entropía y la disponibilidad se desarrollan solo hasta donde exige la sección 6 y corresponden a IDE-003

Trabajos relacionados: IDE-003, IDE-005, IDE-022

Perfil epistémico: Conocimiento establecido en conjunto; la sección 7 lleva una afirmación de Interpretación científica actual

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La energía es una de las magnitudes unificadoras más potentes de la ciencia, pero el lenguaje cotidiano la convierte con facilidad en una sustancia engañosa. Este ensayo examina qué entienden físicos y químicos por energía: una propiedad calculable de un sistema especificado, cuyos cambios se siguen a través del movimiento, la configuración, la radiación, los grados internos de libertad y la masa. Distingue las magnitudes de estado de transferencias como el trabajo y el calor, explica por qué la conservación exige definir bien la frontera del sistema y muestra cómo el teorema de Noether relaciona la energía conservada con la simetría de traslación temporal. Después marca dos límites. Conservar energía no significa que toda esté igualmente disponible para producir cambios útiles, y la relatividad general no ofrece una energía global única para cualquier espaciotiempo. El resultado es un concepto que unifica muchas descripciones sin convertirse en un fluido material ni en una sustancia universal.

En 60 segundos

La energía no es una sustancia escondida dentro de los objetos ni un conjunto de paquetes idénticos que circulan entre ellos. Es una magnitud física que distintas teorías calculan de formas diferentes pero compatibles. Un objeto en movimiento puede aportar energía cinética; la configuración de un sistema, energía potencial; la materia posee energía interna; la radiación transporta energía; y la masa contribuye a la energía relativista. El trabajo y el calor se entienden mejor como maneras en que la energía cruza la frontera de un sistema, no como sustancias almacenadas en su interior. La ley de conservación solo funciona después de especificar el sistema y sus interacciones. La energía mecánica puede disminuir por rozamiento mientras la cuenta energética más amplia sigue equilibrada porque aumenta la energía interna. El teorema de Noether aporta una razón estructural más profunda en muchas teorías físicas: la invariancia frente a traslaciones temporales se asocia a una energía conservada. Pero el lema tiene límites. La entropía restringe la utilidad de la energía y, en un espaciotiempo curvo y evolutivo, puede no existir una única energía total global bien definida. La

energía es, por tanto, un concepto disciplinado de balance con un alcance extraordinario, no un fluido universal.

Por qué importa esta pregunta

La energía aparece en casi todos los relatos científicos: piedras que caen, reacciones químicas, estrellas, baterías, metabolismo y expansión del universo. Esa amplitud invita a tratar la palabra como si nombrara una materia oculta. De ahí nacen confusiones persistentes: el calor parece algo que un objeto «contiene», el rozamiento parece destruir energía, el ATP se imagina como un paquete energético y el corrimiento cosmológico al rojo parece hacer desaparecer energía. Una definición rigurosa hace más que ordenar el vocabulario: obliga a indicar qué se mide, qué frontera se ha elegido, qué magnitud se conserva y cuándo el lema habitual de conservación necesita una condición adicional.

1. Un número que sobrevive al cambio

La energía debe su importancia a que permanece restringida a través de las transformaciones en un sentido preciso y acotado. En mecánica elemental, los términos cinético y potencial pueden intercambiarse. En termodinámica, la energía interna cambia cuando la energía cruza una frontera mediante calor o trabajo. En descripciones electromagnéticas y cuánticas, campos y cuantos transportan energía. La relatividad vincula la masa de un cuerpo con su contenido energético. No son definiciones rivales de sustancias distintas, sino maneras específicas de cada dominio de calcular contribuciones a una estructura común de balance.[1,7-9]

Por eso la lista habitual de «formas de energía» puede inducir a error. La idea útil no es que una materia llamada energía cambie repetidamente de disfraz. La idea útil es que distintas variables de estado y términos dinámicos contribuyen a una magnitud cuyos cambios obedecen relaciones exactas. La famosa analogía de los bloques de Feynman captura el balance, pero también su límite: el número conservado es físicamente significativo aunque no exista una imagen única de cómo «se ven» todas sus contribuciones.[1,7]

El SI refuerza esa abstracción. Energía, trabajo y cantidad de calor se miden en julios, unidad derivada coherente $\text{kg m}^2 \text{s}^{-2}$. [5] Compartir unidad no los convierte en el mismo tipo de concepto. Una magnitud de estado puede describir un sistema en un instante; una magnitud de transferencia describe un proceso a través de una frontera. La ecuación y el procedimiento experimental determinan cuál de las dos se quiere decir, no la unidad por sí sola.

En qué punto estamos

La energía se entiende mejor como una magnitud calculable sometida a restricciones de conservación, no como un ingrediente material. Distintas teorías muestran contribuciones distintas y el julio aporta una unidad común. La siguiente pregunta no es dónde «está realmente» la energía, sino qué sistema se ha definido y qué términos pertenecen a su balance.[1,5,7,8]

2. La frontera del sistema hace el trabajo oculto

Las afirmaciones de conservación siempre se refieren a un sistema. Si dejamos caer una pelota a través del aire, su energía mecánica no permanece exactamente constante: el arrastre reduce la suma de energía cinética y potencial gravitatoria de la pelota. Si ampliamos el sistema para incluir el aire, la pelota y sus grados microscópicos de libertad, la aparente pérdida reaparece sobre todo como energía interna y movimiento en el entorno. No se ha recuperado nada misterioso. Ha cambiado la contabilidad porque cambió la frontera.[8]

Esta es una razón por la que «la energía potencial pertenece al objeto» suele ser una mala frase. Cerca de la Tierra, la energía potencial gravitatoria se asigna de forma natural a la configuración del sistema Tierra-objeto; la energía potencial elástica pertenece a un sistema deformado; la energía potencial electrostática pertenece a una configuración de cargas y campos. En muchos cálculos asignamos términos a un subsistema por conveniencia, pero esa asignación es en parte una elección de modelo. Las predicciones observables permanecen invariantes si la contabilidad se transforma de forma coherente.[14]

La misma disciplina importa en química y biología. Un recipiente de reacción, una célula y un ecosistema intercambian materia y energía con su entorno de maneras diferentes. Antes de afirmar que se produjo, consumió, almacenó o desperdició energía, hay que declarar la frontera y las variables retenidas. Muchas paradojas verbales desaparecen en ese punto.[8]

En qué punto estamos

Un balance energético nunca carece de frontera. Una creación o destrucción aparente suele indicar que la cuenta elegida omitió una transferencia, una interacción o grados internos de libertad. Definir el sistema forma parte de la física y prepara la distinción entre la energía que caracteriza un estado y la que cruza su frontera.[8]

3. Calor y trabajo son transferencias, no fluidos almacenados

La termodinámica afina la distinción. La energía interna es una función de estado: una vez fijado el estado termodinámico, su valor no depende del camino seguido para alcanzarlo. El calor y el trabajo, en cambio, describen transferencias durante un proceso. IUPAC define el calor como energía transferida debido a una diferencia de temperatura, mientras que el trabajo mecánico puede representarse mediante la integral fuerza-desplazamiento.[3,4] Un cuerpo puede tener energía interna; suele ser engañoso decir que «contiene calor» en el mismo sentido técnico.

La primera ley los conecta. Con la convención en la que el trabajo realizado sobre el sistema es positivo, $\Delta U = q + w$. Otras disciplinas usan a veces el signo opuesto para el trabajo, por lo que la convención debe declararse. Lo que no cambia es la estructura: el cambio de la energía almacenada se relaciona con la energía transferida a través de la frontera.[2]

Esto también aclara el rozamiento. No viola la conservación.[8] Redirige energía mecánica organizada hacia movimiento microscópico menos organizado y otros grados internos de libertad. Esa distinción se vuelve crucial cuando preguntamos no solo cuánta energía existe, sino cuánta puede impulsar un cambio especificado.

En qué punto estamos

La energía interna caracteriza un estado; calor y trabajo caracterizan transferencias. Confundirlos crea la falsa imagen del calor como sustancia y hace que el rozamiento parezca destructivo. Al separar estado y transferencia, la primera ley se convierte en una relación de balance y no en un relato sobre fluidos invisibles. Esa distinción prepara la pregunta siguiente: ¿por qué deben conservarse esos balances?[2]

4. La conservación es más profunda que un truco de cálculo

En muchas teorías físicas, la conservación de la energía está vinculada a una simetría. El primer teorema de Noether muestra que las simetrías continuas de una acción corresponden a magnitudes conservadas. Cuando las leyes que gobiernan un sistema cerrado no cambian al desplazar el origen

de la coordenada temporal, la magnitud conservada asociada es la energía. La traslación espacial se relaciona de forma análoga con el momento lineal y la rotación con el momento angular.[6]

Esto no significa que cada ejercicio escolar de conservación requiera resolver un problema variacional. Explica por qué las leyes de conservación están incorporadas estructuralmente en amplias clases de teorías modernas. También evita una exageración frecuente: «la energía se conserva porque existe el tiempo». La condición relevante no es la mera existencia del tiempo, sino una invariancia concreta de la dinámica frente a traslaciones temporales.[6]

El teorema revela además por qué sigue importando definir el sistema. Si una influencia externa explícitamente dependiente del tiempo actúa sobre un subsistema, la energía de ese subsistema puede no conservarse. La teoría más amplia puede poseer una magnitud conservada una vez incluidos los grados de libertad externos, pero la conservación no nos autoriza a ignorar lo que hemos dejado fuera del modelo.

En qué punto estamos

El teorema de Noether relaciona la conservación de la energía con la simetría de traslación temporal en teorías variacionales.[6] Es más preciso que afirmar que la energía se conserva «por definición». También fija el límite: un subsistema abierto o explícitamente dependiente del tiempo puede no poseer por sí solo una energía conservada. El paso siguiente es ver cómo sobrevive esa estructura cuando relatividad y cuantos amplían la cuenta.

5. Masa, luz y ampliación de la cuenta

La relatividad especial amplió la contabilidad clásica. Un cuerpo masivo en reposo posee energía de reposo $E_0 = mc^2$, y su energía relativista total incluye además el movimiento.[1] Esto no significa que la masa sea una sustancia separada que pueda verse en energía. Masa y energía son propiedades relacionadas de un sistema físico,[9] y los cambios en la energía de enlace pueden modificar la masa de un sistema compuesto.

La luz transporta energía aunque los fotones tengan masa en reposo nula. Para un fotón, $E = h\nu$, de modo que la frecuencia está directamente relacionada con la energía. La radiación puede transferir energía y momento, calentar materia, impulsar transiciones y ejercer presión. La imagen clásica según la cual la energía debe pertenecer a partículas masivas es demasiado estrecha.[1,15]

La lección metodológica es que cada ampliación de la física conservó el papel contable de la energía modificando qué términos debían incluirse. La conservación sobrevivió haciéndose más amplia, no forzando fenómenos nuevos dentro de una imagen mecánica antigua.[1,7,9,15]

En qué punto estamos

La relatividad y la física cuántica no abolieron la energía: ampliaron lo que entra en el balance. La masa en reposo contribuye a la energía y la radiación sin masa en reposo la transporta mediante su frecuencia. Lo duradero es la estructura de conservación, no una imagen clásica única de sus términos. Esa ampliación plantea otra pregunta: incluso si la energía total cuadra, ¿cuánta queda disponible para realizar trabajo útil?[1,15]

6. Conservada no significa disponible

Si la energía se conserva, ¿por qué se agota una batería, se enfría una bebida caliente o una máquina pierde capacidad para realizar trabajo? Porque la cantidad total de energía y la capacidad de impulsar un proceso concreto son preguntas distintas.[7] La segunda ley de la termodinámica

restringe el cambio espontáneo mediante la entropía, magnitud cuyo cambio IUPAC define como el calor aportado al sistema en un proceso reversible a temperatura constante, dividido por esa temperatura.[12] Los potenciales como la energía de Gibbs, definida como la entalpía menos el producto de la temperatura termodinámica por la entropía,[13] se construyen para condiciones ambientales especificadas.

Un océano a temperatura ambiente contiene una enorme energía interna, pero un dispositivo en equilibrio con ese mismo océano no puede convertirla toda en trabajo dirigido sin producir otros cambios.[7] A la inversa, una diferencia energética modesta mantenida en forma de gradiente puede ser muy eficaz porque está organizada con respecto al proceso que puede aprovecharla. Este es el punto en el que el lenguaje de «alta energía» y «baja energía» pasa a depender del contexto.

Este ensayo no desarrolla todavía una teoría completa de la entropía. La frontera esencial es que la conservación responde a cómo cierra el balance energético, mientras que la termodinámica pregunta qué transformaciones son posibles, en qué dirección y con qué capacidad de impulso disponible.

En qué punto estamos

La conservación de la energía no garantiza su utilidad. La entropía y las restricciones termodinámicas determinan con qué facilidad las diferencias energéticas pueden impulsar cambios. Un proceso puede conservar perfectamente la energía total y, al mismo tiempo, reducir irreversiblemente la fracción convertible en el trabajo deseado. El último límite es geométrico: ¿qué ocurre con el lema cuando el propio espaciotiempo es dinámico? [7]

7. El límite cosmológico del lema

La precaución más fuerte aparece en relatividad general. La materia se describe mediante un tensor de energía-momento cuya divergencia covariante se anula, lo que expresa una estructura local de conservación. Pero el espaciotiempo curvo complica la idea de una densidad única de energía gravitatoria, y no existe en todo espaciotiempo una única energía total global conservada. Las definiciones globales funcionan limpiamente solo bajo condiciones geométricas adicionales, como simetrías adecuadas o determinada estructura asintótica.[10,11]

Esto importa en cosmología. En un universo de Friedmann-Lemaître-Robertson-Walker en expansión, las longitudes de onda de los fotones se estiran y sus energías medidas disminuyen. Preguntar «¿adónde fue esa energía?» supone de antemano que el universo evolutivo completo posee el mismo tipo de escalar global conservado familiar en sistemas de laboratorio invariantes bajo traslaciones temporales. La relatividad general no ofrece ese objeto en general.[10,11] Las ecuaciones locales siguen siendo coherentes; lo que se ha llevado más allá de su dominio es la intuición del balance global.

No es una prueba de que la conservación sea aproximada u opcional. Es una prueba de que las leyes de conservación poseen condiciones matemáticas. Un lema aprendido en sistemas aislados y en un espaciotiempo casi plano no debe elevarse a un teorema más fuerte que el que la propia teoría sostiene.[10,11]

En qué punto estamos

La relatividad general conserva una estructura local de energía-momento, pero no garantiza una energía global única y conservada para cualquier espaciotiempo.[10,11] Por eso el corrimiento cosmológico al rojo no debe interpretarse exigiendo un depósito global de energía desaparecida. La

condición de contorno forma parte de la ley, no es un estorbo añadido después. Ese límite prepara la síntesis final: separar el balance robusto de las condiciones que permiten formular una conservación global.

Lo que sabemos y cómo lo sabemos

- La energía es una propiedad definida cuantitativamente y utilizada en mecánica, termodinámica, electromagnetismo, física cuántica y relatividad.[1,5,7]
- Experimentos y teoría respaldan relaciones exactas de balance cuando se especifican el sistema y las transferencias relevantes.[8]
- Calor y trabajo son magnitudes de proceso; la energía interna es una magnitud de estado.
- El teorema de Noether conecta la invariancia bajo traslaciones temporales con una energía conservada en teorías variacionales.[6]
- La relatividad especial exige incluir la energía de reposo y las contribuciones relativistas, y la radiación transporta energía.[1,15]
- La conservación por sí sola no determina cuánta energía está disponible para realizar trabajo útil.[7]
- La relatividad general conserva localmente energía-momento, pero no proporciona una energía global única en todo espaciotiempo.[10,11]

Lo que todavía no sabemos

- Cómo definir de la forma más útil la energía gravitatoria cuasilocal en relatividad general, y cuál de las definiciones disponibles conviene en un espaciotiempo concreto.[10,11]
- Cómo conectarán las futuras teorías de gravedad cuántica sus magnitudes conservadas con las descripciones de baja energía ya familiares.
- Nada de esto es una duda empírica abierta sobre el funcionamiento de los balances energéticos ordinarios de sistemas cerrados dentro de sus dominios probados. Estas preguntas no debilitan las leyes de conservación de laboratorio; marcan el punto donde el concepto se encuentra con la estructura del propio espaciotiempo.[10,11]

Qué podría hacernos cambiar de opinión

- Un fallo reproducible del balance energético en un sistema genuinamente aislado, después de incluir todas las partículas, campos, radiación y transferencias de frontera conocidas, exigiría una revisión profunda.[6]
- También lo haría demostrar que una teoría fundamental invariante bajo traslaciones temporales carece sistemáticamente de la magnitud conservada correspondiente cuando se cumplen los supuestos de Noether.[6]
- En gravitación, un nuevo marco fundamental podría cambiar qué definiciones globales o cuasilocales de energía se prefieren, pero tendría que recuperar los límites locales y de campo débil ya medidos con enorme éxito.[10,11]

Si solo recuerdas una cosa

La energía es una magnitud de balance cuyo significado depende del sistema y la teoría; no es una sustancia, y su conservación global tiene límites geométricos.

Artículos relacionados

- IDE-003 — Entropía
- IDE-005 — Simetría
- IDE-022 — Masa

Referencias

- [1] International Union of Pure and Applied Chemistry. Energy. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.E02101.
- [2] International Union of Pure and Applied Chemistry. Internal energy. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.I03103.
- [3] International Union of Pure and Applied Chemistry. Heat. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.H02752.
- [4] International Union of Pure and Applied Chemistry. Work. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.W06684.
- [5] Bureau International des Poids et Mesures. The International System of Units (SI), 9th edition, updated 2026. DOI 10.59161/AUEZ1291.
- [6] Emmy Noether. Invariant Variation Problems. English translation by M. A. Tavel. Transport Theory and Statistical Physics 1, 186-207 (1971). DOI 10.1080/00411457108231446.
- [7] Richard P. Feynman; Robert B. Leighton; Matthew Sands. The Feynman Lectures on Physics, New Millennium Edition, Vol. I, chapter 4 (Conservation of Energy). Basic Books, ISBN 978-0-465-02414-8.
- [8] William Moebs; Samuel J. Ling; Jeff Sanny. University Physics Volume 1, section 8.3 (conservation of energy). OpenStax (2016).
- [9] Albert Einstein. Does the Inertia of a Body Depend Upon Its Energy-Content? Annalen der Physik 18, 639-641 (1905). DOI 10.1002/andp.19053231314.
- [10] László B. Szabados. Quasi-Local Energy-Momentum and Angular Momentum in General Relativity. Living Reviews in Relativity 12, 4 (2009). DOI 10.12942/lrr-2009-4.
- [11] Patrick M. Dürr. Fantastic Beasts and where (not) to find them: Local gravitational energy and energy conservation in general relativity. Studies in History and Philosophy of Modern Physics 65, 1-14 (2019). DOI 10.1016/j.shpsb.2018.07.002.
- [12] International Union of Pure and Applied Chemistry. Entropy. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.E02149.
- [13] International Union of Pure and Applied Chemistry. Gibbs energy. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.G02629.
- [14] William Moebs; Samuel J. Ling; Jeff Sanny. University Physics Volume 1, section 8.1 (potential energy of a system). OpenStax (2016).
- [15] International Union of Pure and Applied Chemistry. Photon. Gold Book, 5th ed. (2025). DOI 10.1351/goldbook.P04627.

Información: cuantificar alternativas y dependencia

IDE-002 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.6 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Información estadística y física a través de escalas

Límite: Información cuantitativa y codificación; las señales físicas y la señalización se tratan en IDE-024

Trabajos relacionados: IDE-003, IDE-008, IDE-024

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La información es un lenguaje cuantitativo para alternativas distinguibles, incertidumbre estadística y dependencia. En la teoría de Shannon, una fuente se representa mediante mensajes posibles con probabilidades; la entropía resume la incertidumbre media, mientras la información mutua mide cuánto reduce la observación de una variable la incertidumbre sobre otra. Estas magnitudes fijan límites rigurosos para compresión y comunicación fiable, pero por sí solas no aportan significado semántico, dirección causal ni una sustancia física llamada información. Un bit es una unidad definida por logaritmos en base dos, y la misma distinción lógica puede encarnarse en muchos sistemas físicos. La implementación física, sin embargo, importa: la memoria debe representarse, manipularse y a veces borrarse, y las operaciones lógicamente irreversibles tienen consecuencias termodinámicas. Este artículo separa la medida matemática de sus soportes físicos y de la señalización causal, fijando la frontera con IDE-024, que pregunta qué es una señal y cómo se transmite físicamente una influencia.

En 60 segundos

Empieza con un conjunto de resultados posibles y sus probabilidades. La entropía de Shannon, medida en bits cuando se usan logaritmos en base dos, da la incertidumbre media de esa distribución; la información mutua mide dependencia estadística entre dos variables. No son metáforas vagas. Sustentan teoremas de codificación: los resultados frecuentes pueden recibir descripciones más cortas, los canales ruidosos tienen capacidades finitas y procesar datos no puede crear información sobre una variable anterior que no estuviera en la entrada.[1,2] Pero la matemática abstrae deliberadamente el significado. Dos cadenas de bits pueden tener las mismas estadísticas de Shannon y diferir por completo en lo que una persona interpreta que significan.[1] La información tampoco es un fluido sin peso independiente de su representación. Un bit lógico puede almacenarse mediante voltaje, magnetización u otro grado de libertad físico, mientras borrar estados lógicamente distinguibles bajo condiciones térmicas conlleva un coste termodinámico mínimo en el límite ideal.[3,4] Por último, información compartida no demuestra comunicación ni causalidad: causas comunes e interacciones previas pueden correlacionar variables sin una señal presente entre ellas.[5]

Por qué importa esta pregunta

La palabra información viaja con facilidad entre telecomunicaciones, física, biología, computación y lenguaje cotidiano, y esa portabilidad es útil y peligrosa a la vez. Sin declarar variables aleatorias, probabilidades y supuestos de codificación, una afirmación numérica sobre información puede quedar indefinida; sin un modelo causal, la información mutua puede confundirse con influencia; sin un modelo físico, el principio de Landauer puede citarse erróneamente como coste energético universal de cada paso computacional. Un tratamiento disciplinado permite usar las mismas herramientas matemáticas entre campos manteniendo separadas preguntas distintas: cuánta incertidumbre tiene una fuente, cuánto dependen dos variables, cómo se codifica un mensaje y cómo viaja realmente una señal física.

1. Qué mide la información de Shannon

El punto de partida de Shannon no es una sustancia, sino un conjunto de mensajes posibles y sus probabilidades. Para una variable aleatoria discreta X con resultados x y probabilidades $p(x)$, la entropía $H(X) = -\sum p(x) \log_2 p(x)$ es la sorpresa logarítmica esperada. La base del logaritmo fija la unidad: base dos da bits. La entropía es cero cuando un resultado es seguro y aumenta cuando la probabilidad se reparte entre alternativas.[1,2]

La definición es operacional porque conecta con la codificación. Para secuencias largas generadas por una fuente, la entropía fija la escala fundamental de la longitud media de descripción sin pérdida: las fuentes muy predecibles pueden comprimirse más que las muy inciertas. La teoría de Shannon para canales con ruido relaciona igualmente la comunicación fiable con la capacidad del canal y no con un alfabeto o hardware particulares.[1,2] La abstracción es potente precisamente porque ignora la forma material de los símbolos.

La información mutua plantea otra pregunta. $I(X;Y)$ cuantifica cuánto se aparta de la independencia la distribución conjunta de X e Y , o de forma equivalente cuánto reduce conocer una la incertidumbre sobre la otra. Es no negativa y se anula cuando las variables son estadísticamente independientes bajo la distribución especificada.[2] Un valor alto significa, por tanto, dependencia estadística fuerte, no que X haya causado Y ni que un mensaje haya viajado físicamente de X a Y .

En qué punto estamos

La teoría de la información empieza con alternativas explícitas y distribuciones de probabilidad y las convierte en magnitudes que restringen compresión y comunicación. Entropía e información mutua son propiedades de una descripción probabilística, no fluidos ocultos dentro de objetos. Esto hace inevitable la siguiente pregunta: si la matemática abstrae el hardware, ¿qué cambia cuando las mismas distinciones lógicas se implementan en dispositivos físicos reales?

2. Un bit es lógico, pero su implementación es física

Un bit es una distinción binaria, no una partícula microscópica. Los estados lógicos 0 y 1 pueden representarse por dos rangos de voltaje, dos direcciones de magnetización, dos estados ópticos o muchas otras configuraciones físicamente distinguibles. La teoría de la información puede tratar estas realizaciones como equivalentes cuando solo importan las alternativas lógicas, mientras la ingeniería debe seguir además ruido, estabilidad, barreras energéticas, probabilidades de error y dinámica de lectura o cambio del estado.[1,2]

Landauer identificó un puente concreto entre irreversibilidad lógica y termodinámica. Cuando una operación física hace corresponder dos estados lógicos distinguibles con uno solo, como ocurre en

el borrado idealizado, se descarta información sobre el estado previo. En un entorno térmico, realizar esa operación lógicamente irreversible de la forma más reversible posible exige aun así exportar entropía al entorno; para borrar un bit imparcial, el límite inferior ideal familiar es $kT \ln 2$ de calor disipado.[3]

El principio no afirma que toda operación lógica deba disipar $kT \ln 2$. Bennett mostró que, en principio, la computación general puede organizarse mediante pasos lógicamente reversibles, conservando información suficiente para recorrer hacia atrás estados intermedios y evitando el borrado hasta que sea realmente necesario.[4] Los ordenadores reales disipan mucho más por numerosas razones de ingeniería, pero la lección conceptual es precisa: la información es suficientemente abstracta para admitir muchas encarnaciones, mientras las operaciones sobre información encarnada siguen sujetas a la ley física.

En qué punto estamos

El mismo bit lógico puede habitar hardware muy distinto, por lo que la información no puede identificarse con un soporte material único. Sin embargo, la implementación no es irrelevante: almacenar, leer y borrar distinciones son procesos físicos con ruido y restricciones termodinámicas. Esto separa la información como relación cuantitativa entre alternativas de la señal que transporta físicamente un cambio entre lugares, tema reservado a IDE-024.

3. Información no es lo mismo que significado

Shannon diseñó explícitamente su teoría de la comunicación de modo que los aspectos semánticos de los mensajes no fueran necesarios para el problema de ingeniería.[1] Un código puede conservar una secuencia con error arbitrariamente bajo sin comprender si esa secuencia es un poema, un genoma, un archivo de software o una salida de apariencia aleatoria. La matemática atiende a probabilidades de símbolos, longitudes de código, estadísticas del canal y recuperabilidad bajo una tarea especificada.

Esto no afirma que el significado sea irreal. Es una frontera sobre lo que miden las magnitudes de Shannon. El significado depende de usuarios, convenciones, referencia, objetivos y modelos más amplios que $H(X)$ por sí sola no determina. Dos mensajes pueden tener longitud y entropía similares y diferir radicalmente en contenido semántico; a la inversa, la misma proposición puede codificarse en muchas cadenas con propiedades estadísticas distintas. Las medidas cuantitativas de información no se convierten, por compartir una palabra, en una teoría universal de la semántica.

El contexto entra, no obstante, en la matemática de forma disciplinada. Hay que elegir variables, un alfabeto o espacio de estados, un modelo probabilístico y a menudo un protocolo de codificación u observación antes de calcular entropía o información mutua. Elecciones distintas pueden responder preguntas legítimas distintas. Una vez fijadas, sin embargo, la magnitud resultante no es arbitraria: el modelo probabilístico y las definiciones determinan el valor y los teoremas restringen qué pueden hacer las transformaciones.[2]

En qué punto estamos

La información de Shannon responde deliberadamente a una pregunta más estrecha que el significado del lenguaje ordinario. Cuantifica incertidumbre y dependencia una vez especificadas variables y probabilidades; no infiere a qué se refieren los símbolos ni por qué importan a un organismo u observador. Esa frontera evita un error de categoría frecuente y prepara el siguiente: puede haber dependencia estadística fuerte sin influencia causal ni comunicación presente.

4. Dependencia no es causalidad ni señalización

La información mutua detecta dependencia estadística. Si dos variables comparten una causa común, quedaron correlacionadas por una interacción anterior o son ambas efectos de un tercer proceso, observar una puede reducir incertidumbre sobre la otra aunque ninguna envíe actualmente nada a la otra. La teoría de la información por sí sola no puede decidir si la dependencia surgió por influencia directa, causa común, selección u otra estructura causal.[2,5]

La inferencia causal requiere supuestos adicionales o intervenciones. El marco gráfico de Pearl lo hace explícito al combinar distribuciones de probabilidad con información estructural sobre qué variables pueden afectar directamente a cuáles; solo bajo tales supuestos los datos observacionales pueden identificar efectos causales, y a veces no pueden hacerlo.[5] Esta es una tarea distinta de medir dependencia. La diferencia importa siempre que frases como «la información fluye de A a B» se usen como si una correlación no nula ya hubiera demostrado una ruta física.

La desigualdad de procesamiento de datos aporta una restricción útil pero limitada. Si X influye en Y y después Y se procesa para formar Z mediante una cadena de Markov $X \rightarrow Y \rightarrow Z$, el procesamiento ordinario no puede aumentar la información mutua que Z contiene sobre X más allá de la que ya contenía Y.[2] El teorema restringe la estructura estadística dado un modelo; no establece que la cadena causal asumida sea la historia física verdadera. Modelo y medida deben mantenerse separados.

En qué punto estamos

La información compartida puede surgir sin una señal contemporánea, de modo que dependencia, causalidad y comunicación no deben colapsarse en una sola palabra. La teoría de la información restringe distribuciones de probabilidad; la inferencia causal añade supuestos sobre intervención y dirección; la física de señales pregunta qué interacción puede transportar realmente un cambio. Separados esos papeles, queda precisar qué unifica la información sin convertirla en sustancia universal.

5. Qué unifica el concepto y dónde se detiene

La teoría de la información unifica problemas que pueden expresarse como alternativas, probabilidades, restricciones de codificación y dependencia estadística. Las mismas herramientas formales pueden describir compresión de fuentes, comunicación ruidosa, inferencia y partes de la física estadística porque cada problema contiene una estructura de estados distinguibles e incertidumbre.[1,2] Esa matemática común es sustantiva, pero no implica que cada uso de la palabra información nombre el mismo mecanismo físico.

El concepto tiene también fronteras dentro de la matemática. La entropía diferencial puede ser negativa y cambia bajo transformaciones de coordenadas y cambios de escala, mientras que la información mutua tiene propiedades de invariancia más fuertes.[2] Afirmaciones como «el sistema contiene N bits» requieren por tanto especificar representación, resolución y modelo probabilístico.

La división del trabajo en esta serie es deliberada. IDE-003 trata la entropía como concepto más amplio, incluido su papel termodinámico; IDE-008 trata la probabilidad en sí; IDE-024 pregunta qué es una señal, centrándose en soportes físicos, propagación y condiciones de señalización. IDE-002 se detiene en la información cuantitativa: cómo se representan y restringen alternativas,

incertidumbre, dependencia y codificación. El relevo es conceptual, no probatorio, y por ello no necesita cita.

En qué punto estamos

La información es una familia rigurosa de relaciones cuantitativas, no un ingrediente material universal. Su potencia procede de separar lo que depende de probabilidad y codificación de lo que exige semántica, estructura causal o un mecanismo físico de transmisión. Esa frontera hace complementarios, y no redundantes, los artículos vecinos: entropía, probabilidad y señal heredan cada uno parte de la historia, pero ninguno es simplemente otro nombre para información.

Lo que sabemos y cómo lo sabemos

- Para fuentes discretas, la entropía de Shannon cuantifica incertidumbre media y fija la escala fundamental de la compresión sin pérdida.[1,2]
- La información mutua mide dependencia estadística y obedece la desigualdad de procesamiento de datos bajo la estructura probabilística especificada.[2]
- El borrado lógicamente irreversible tiene un límite termodinámico inferior en el límite térmico ideal, mientras la computación lógicamente reversible puede en principio evitar pagar ese coste en cada paso.[3,4]
- La dependencia estadística por sí sola no identifica dirección causal; las afirmaciones causales requieren supuestos estructurales adicionales o intervenciones.[5]

Lo que todavía no sabemos

- Cómo conectar mejor magnitudes de teoría de la información con nociones semánticas o funcionales en sistemas biológicos y cognitivos concretos sin confundir definiciones.
- Qué medidas de información resultan más informativas para sistemas muy no estacionarios y de alta dimensión cuando las propias variables de estado relevantes son inciertas.
- Hasta qué punto futuras arquitecturas físicas de computación podrán acercarse a los límites de computación reversible sin perder velocidad, fiabilidad y fabricabilidad.

Qué podría hacernos cambiar de opinión

- Una violación reproducible de los límites de codificación de fuente sin pérdida o de codificación de canal bajo sus supuestos declarados obligaría a revisar el marco correspondiente de teoría de la información.
- Un experimento controlado de borrado que, tras contabilizar todos los reservorios y errores, disipara repetidamente menos que el límite de Landauer para un bit imparcial en el régimen térmico cuasiestático pondría en cuestión la explicación termodinámica estándar.[3]
- Un procedimiento de inferencia causal que identificara de forma única la dirección a partir de distribuciones conjuntas observacionales arbitrarias sin supuestos adicionales, y funcionara también con causas comunes y equivalencia observacional, derribaría la frontera aquí trazada entre dependencia y causalidad.[5]

Si solo recuerdas una cosa

La información cuantifica incertidumbre y dependencia estadística bajo elecciones explícitas de variables y probabilidades; no es significado semántico, influencia causal ni una sustancia física.

Artículos relacionados

- IDE-003 — Entropía
- IDE-008 — Probabilidad
- IDE-024 — ¿Qué es una señal?

Referencias

- [1] Claude E. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423 and 623-656 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x; 10.1002/j.1538-7305.1948.tb00917.x.
- [2] Thomas M. Cover; Joy A. Thomas. Elements of Information Theory, 2nd ed. Wiley (2006). DOI 10.1002/047174882X.
- [3] Rolf Landauer. Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development 5(3), 183-191 (1961). DOI 10.1147/rd.53.0183.
- [4] Charles H. Bennett. Logical Reversibility of Computation. IBM Journal of Research and Development 17(6), 525-532 (1973). DOI 10.1147/rd.176.0525.
- [5] Judea Pearl. Causal diagrams for empirical research. Biometrika 82(4), 669-688 (1995). DOI 10.1093/biomet/82.4.669.

Entropía: conteo, restricciones e irreversibilidad

IDE-003 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.8 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Termodinámica, mecánica estadística y teoría de la información

Límite: Se comparan entropía termodinámica, entropía estadística y entropía de Shannon sin tratarlas como intercambiables

Trabajos relacionados: IDE-002, IDE-010, IDE-017

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La entropía aparece en termodinámica, mecánica estadística y teoría de la información, pero la misma palabra no designa una única magnitud sin restricciones. En termodinámica de equilibrio es una función de estado que entra en la segunda ley.[4] La mecánica estadística la conecta con distribuciones de probabilidad sobre posibilidades microscópicas compatibles con restricciones declaradas.[1,2] La entropía de Shannon cuantifica la incertidumbre en las probabilidades de los mensajes y comparte la forma $-p \log p$ utilizada en mecánica estadística.[3] Estos usos se solapan matemáticamente pero difieren en variables, unidades y papel operacional. El artículo explica por qué la entropía local puede disminuir, por qué importa el coarse-graining, por qué el desorden visible no es una definición y cómo el borrado de información crea un puente termodinámico específico.[5,6] El resultado es una imagen controlada de la entropía como contabilidad cuantitativa cuya interpretación depende del sistema, ensemble, restricciones y proceso físico considerados.

En 60 segundos

La entropía termodinámica es una variable de estado. Para un sistema macroscópico aislado, la segunda ley establece que la evolución espontánea no reduce la entropía total; el equilibrio corresponde al máximo de entropía compatible con las cantidades conservadas y restricciones impuestas.[4] La mecánica estadística aporta un lenguaje microscópico para esa regla. Un macroestado puede ser compatible con muchos microestados, y las distribuciones de probabilidad sobre estados microscópicos conducen a la entropía de Gibbs.[1,2] Shannon introdujo después una entropía para la incertidumbre en conjuntos de mensajes; matemáticamente presenta la misma forma $-p \log p$, pero las variables y la interpretación son distintas.[3] Por eso «entropía es desorden» es una mala definición: una baraja mezclada, un cristal, un gas y un mensaje codificado requieren espacios de estados y restricciones diferentes, y el aspecto visual no dice qué entropía se calcula.[5] La entropía local puede disminuir cuando se exporta entropía a otra parte: los frigoríficos enfrían compartimentos, los cristales forman estructuras ordenadas y los seres vivos mantienen organización mientras producen entropía en su entorno.[4] El borrado irreversible de información ofrece un puente concreto entre distinciones lógicas y

producción de entropía termodinámica, pero no una licencia para identificar toda entropía informacional con calor.[6]

Por qué importa esta pregunta

La entropía aparece en física, química, computación, cosmología y explicaciones populares del tiempo. Esa amplitud vuelve especialmente costosos los traslados descuidados. Si se trata como una medida mística del caos, la organización local puede parecer una violación de la segunda ley; si la entropía de Shannon se vuelve automáticamente termodinámica, un cálculo de probabilidad puede confundirse con un balance de calor. La disciplina científica consiste en nombrar sistema, variables de estado, ensemble, frontera e intercambios permitidos antes de concluir. Con esas piezas explícitas, la entropía se vuelve una herramienta potente para irreversibilidad e incertidumbre en lugar de un eslogan que pretende explicarlo todo.

1. La entropía termodinámica es una función de estado

La termodinámica clásica define la entropía mediante relaciones entre estados de equilibrio, transferencia de calor y temperatura, no mediante una imagen de desorden molecular. Para una transferencia reversible de calor, la relación diferencial $dS = \delta Q_{\text{rev}}/T$ permite calcular diferencias de entropía, mientras la segunda ley limita los procesos reales al exigir que el cambio total de entropía de un sistema aislado sea no negativo.[4] La entropía queda así ligada a un sistema especificado y a sus variables de estado.

La palabra total es esencial. Un subsistema puede perder entropía mientras su entorno gana más. Un frigorífico reduce la entropía asociada al compartimento frío consumiendo trabajo y expulsando calor a la habitación; el proceso completo dispositivo-entorno respeta la segunda ley.[4] La misma lógica vale cuando crece un cristal desde una disolución o cuando un organismo mantiene organización interna. El orden local no abre un resquicio porque la frontera de contabilidad no puede elegirse después del resultado.

En equilibrio, el principio de entropía puede expresarse como un máximo sujeto a las restricciones relevantes. Esas restricciones pueden incluir energía, volumen, número de partículas o condiciones externas impuestas.[4] Si cambian las restricciones, cambia la comparación. La entropía no ordena todos los estados imaginables en una única escala cósmica; es una función de estado usada dentro de una descripción termodinámica donde variables e intercambios están definidos.

En qué punto estamos

La entropía termodinámica es rigurosa porque la frontera del sistema y sus restricciones forman parte de la afirmación. La segunda ley trata el balance completo, por lo que permite disminuciones locales compensadas en otro lugar. Esa regla macroscópica deja una pregunta más profunda: ¿por qué dominan los estados de equilibrio y cómo puede surgir una magnitud definida sin desorden visible a partir de posibilidades microscópicas? La mecánica estadística construye ese puente.

2. La mecánica estadística cuenta posibilidades bajo restricciones

La mecánica estadística representa un estado macroscópico mediante un ensemble de posibilidades microscópicas. Gibbs formuló una entropía para una distribución de probabilidad sobre el espacio de fases, $S = -k \sum p_i \ln p_i$ en una representación discreta, con versiones continuas que requieren cuidados adicionales.[1] Cuando los microestados accesibles reciben igual probabilidad, la expresión se reduce a la dependencia logarítmica familiar de Boltzmann con el número de

microestados compatibles. El logaritmo hace aditivas las multiplicidades de sistemas independientes.

La frase importante es compatibles con las restricciones macroscópicas. Un gas a energía y volumen fijos no explora configuraciones arbitrarias con la misma relevancia; el ensemble queda definido por lo que se mantiene fijo y por los supuestos dinámicos y probabilísticos del modelo.[1,2] Jaynes subrayó que la mecánica estadística puede entenderse como inferencia con información incompleta: maximizar la entropía sujeto a valores esperados conocidos en vez de inventar detalle microscópico no respaldado.[2] Esa interpretación es potente, pero no elimina el contenido físico de las restricciones.

El coarse-graining aparece cuando se ignoran muchas distinciones microscópicas a favor de variables macroscópicas. Particiones diferentes del espacio de fases pueden producir entropías coarse-grained distintas, y la evolución hamiltoniana de grano fino presenta sutilezas que ocultan los eslóganes de «la entropía siempre aumenta microscópicamente».[1,2] La flecha termodinámica observada depende de la preparación, la resolución macroscópica y el peso abrumador de regiones de espacio de fases asociadas a macroestados de equilibrio, no de que las moléculas se vuelvan literalmente más desordenadas.

En qué punto estamos

La entropía estadística conecta la irreversibilidad macroscópica con alternativas microscópicas ponderadas, pero solo después de elegir ensemble y restricciones. El coarse-graining no es un adorno verbal: especifica qué distinciones retiene la descripción macroscópica. La misma matemática de alternativas ponderadas aparece en teoría de la comunicación, aunque allí se cuentan mensajes y no estados moleculares. Esa semejanza exige fijar una frontera cuidadosa.

3. Entropía de Shannon y entropía estadística: forma compartida, condiciones distintas

La entropía de Shannon $H = -\sum p_i \log p_i$ cuantifica incertidumbre en una distribución de probabilidad sobre símbolos o mensajes.[3] Su semejanza formal con la entropía de Gibbs basta para sostener conexiones profundas, y Jaynes utilizó deliberadamente razonamiento de teoría de la información para construir ensembles estadísticos.[2] Pero unidad, variables y tarea operacional difieren. La entropía de Shannon en bits pertenece a un problema de codificación; la entropía termodinámica en julios por kelvin participa en balances de energía y calor. Inferencia: la notación compartida no establece por sí sola identidad termodinámica; esa identificación requiere las restricciones y los supuestos de ensemble pertinentes.

La distinción evita dos errores opuestos. Primero, a un sistema físico pueden asignársele entropías de Shannon para muchas variables elegidas, ninguna de las cuales tiene por qué coincidir con su entropía termodinámica. Segundo, puede usarse entropía termodinámica sin afirmar que la naturaleza esté literalmente procesando mensajes. La matemática compartida suele revelar una estructura común - probabilidad sobre alternativas y aditividad logarítmica - mientras la interpretación procede del problema físico o informacional en el que se inserta la fórmula.[2,3]

El principio de Landauer muestra un puente genuino pero específico. Reiniciar un bit lógicamente desconocido a un estado estándar es una operación lógica muchos-a-uno. Cuando se implementa físicamente en un entorno térmico, la distinción lógica perdida debe acompañarse de producción

de entropía en otra parte, dando la escala ideal $kT \ln 2$ para borrar un bit imparcial.[6] Esto conecta tratamiento de información y termodinámica mediante una operación y una implementación física; no convierte cada bit de Shannon en un paquete fijo de calor.

En qué punto estamos

Las entropías de Shannon y termodinámica pertenecen a una familia matemática común, pero su igualdad nunca queda garantizada por la notación. Un puente válido requiere codificación física, distribución de probabilidad y proceso definidos, como ilustra el borrado de Landauer. Una vez asegurada esa distinción, puede examinarse directamente el atajo popular más persistente: por qué el desorden visual no es ni definición ni medición fiable de la entropía.

4. Por qué «entropía significa desorden» falla

La metáfora del desorden atrae porque los gases se expanden, el hielo se funde y las mezclas suelen parecer menos estructuradas cuando aumenta la entropía. Pero la apariencia no es la variable de las ecuaciones. Lambert argumentó que «desorden» es una mala muleta docente porque anima a inferir entropía desde el aspecto de desorganización en vez de desde dispersión de energía, multiplicidad y el proceso termodinámico especificado.[5] Una baraja mezclada parece aleatoria, pero barajar cartas no constituye una medida de entropía termodinámica.

Ejemplos visualmente ordenados pueden poseer entropía alta bajo una descripción, mientras estados visualmente irregulares pueden estar fuertemente restringidos bajo otra. Un cristal a temperatura no nula contiene grados vibracionales; un vidrio puede tener complejidad configuracional; una distribución de probabilidad sobre una secuencia de símbolos perfectamente regular puede tener baja entropía de Shannon aunque el soporte físico esté caliente. La pregunta correcta no es «¿parece ordenado?», sino «¿qué estados, probabilidades y restricciones definen la entropía discutida?»[1-5]

La metáfora también oculta la escala. Mezclar dos gases ideales, dejar fluir calor a través de una diferencia de temperatura o borrar un bit son procesos distintos con balances de entropía diferentes. Llamarlos a todos «más desorden» esconde el mecanismo y puede volver circular la segunda ley. La entropía solo adquiere poder explicativo cuando se vincula a un modelo cuantitativo que predice un cambio o restringe un proceso permitido.

En qué punto estamos

El desorden visible puede acompañar algunos aumentos de entropía, pero no puede definirla porque no especifica espacio de estados, distribución de probabilidad ni frontera termodinámica. Sustituir el eslogan por variables explícitas también aclara cómo la entropía puede sostener una flecha temporal sin ser un reloj universal. El último paso separa la asimetría estadística de los procesos macroscópicos irreversibles de las leyes microscópicas usadas para modelarlos.

5. Entropía, irreversibilidad y flecha del tiempo

Muchas leyes dinámicas microscópicas son aproximadamente simétricas bajo inversión temporal, mientras procesos macroscópicos como difusión, fricción y flujo de calor muestran una dirección fuerte en condiciones ordinarias. La mecánica estadística reconcilia ambos hechos combinando dinámica con preparaciones especiales de baja entropía, descripciones macroscópicas coarse-grained y el peso enorme de los macroestados cercanos al equilibrio en el espacio de fases.[1,2] La

segunda ley es estadística en su interpretación microscópica y, a la vez, extraordinariamente fiable para sistemas macroscópicos.

Esa fiabilidad no equivale a imposibilidad lógica de toda fluctuación. Sistemas pequeños pueden mostrar fluctuaciones transitorias que reducen entropía bajo descripciones estocásticas cuidadosamente definidas, y las relaciones de fluctuación cuantifican esos sucesos. Para la termodinámica ordinaria importa la escala: fluctuaciones capaces de deshacer un aumento macroscópico de entropía son extraordinariamente improbables. La flecha macroscópica surge de probabilidad, restricciones y condiciones iniciales, no de una nueva fuerza microscópica que empuje el tiempo hacia delante.

La entropía tampoco explica por sí sola por qué el universo comenzó en una condición de baja entropía ni resuelve todas las preguntas sobre flechas cosmológicas del tiempo. Esos son problemas más profundos de condiciones de frontera. Dentro de su dominio establecido, sin embargo, la entropía ofrece un lenguaje preciso para separar idealizaciones reversibles de procesos irreversibles y rastrear dónde debe ir la entropía compensatoria. IDE-010 tratará la retroalimentación e IDE-017 la incertidumbre; ninguno sustituye esta contabilidad termodinámica.

En qué punto estamos

La entropía enlaza dirección termodinámica, multiplicidad estadística e incertidumbre informacional sin reducirlas a un único significado. Su fuerza explicativa procede de restricciones, ensembles y fronteras explícitas. El concepto sobrevive así tanto al orden local como a la reversibilidad microscópica: la irreversibilidad pertenece al comportamiento estadístico de procesos físicos bajo condiciones especificadas, no a una tendencia visual hacia el desorden. Ese es el significado controlado que continúa en la serie.

Lo que sabemos y cómo lo sabemos

- Para un sistema macroscópico aislado, la segunda ley exige producción total de entropía no negativa; disminuciones locales son compatibles cuando quedan compensadas por cambios mayores en otra parte.[4]
- La entropía de Gibbs y otras entropías estadísticas cuantifican alternativas microscópicas ponderadas bajo ensembles y restricciones especificados.[1,2]
- La entropía de Shannon cuantifica la incertidumbre en las probabilidades de los mensajes y comparte la forma $-p \log p$ utilizada en mecánica estadística.[2,3]
- La metáfora del desorden no define la entropía y puede clasificar mal procesos cuando se omiten variables de estado, fronteras o distribuciones de probabilidad.[5]
- El límite de Landauer conecta el borrado lógicamente irreversible con producción de entropía termodinámica para una implementación física en un entorno térmico.[6]

Lo que todavía no sabemos

- Cómo explicar en última instancia la condición de baja entropía del universo temprano, en vez de limitarse a usarla como condición inicial en las explicaciones estadísticas de la flecha termodinámica.
- Cómo definir y medir mejor la producción de entropía en sistemas fuertemente fuera del equilibrio cuando no existe una descripción termodinámica simple del estado.

- Qué variables coarse-grained ofrecen la descripción macroscópica más informativa y robusta en sistemas complejos cuyos grados de libertad relevantes cambian con el tiempo.

Qué podría hacernos cambiar de opinión

- Un experimento reproducible con un sistema macroscópico aislado que mostrara producción total de entropía negativa sostenida después de cerrar cuantitativamente todos los reservorios de calor, trabajo, materia e información contradiría la formulación estándar de la segunda ley. [4]
- Un experimento de codificación que violara sistemáticamente los teoremas de fuente o canal de Shannon bajo sus supuestos probabilísticos declarados obligaría a revisar el marco de entropía informacional y no solo su interpretación física.[3]
- Un experimento controlado de borrado de un bit en régimen térmico cuasiestático que, tras contabilizar completamente los reservorios, disipara repetidamente por debajo del límite de Landauer cuestionaría el vínculo estándar entre irreversibilidad lógica y producción de entropía termodinámica.[6]

Si solo recuerdas una cosa

La entropía cuantifica alternativas restringidas o cambios de estado termodinámicos bajo una descripción especificada; no es una medida universal del desorden visible.

Artículos relacionados

- IDE-002 — Información
- IDE-010 — Retroalimentación
- IDE-017 — Incertidumbre

Referencias

- [1] Josiah Willard Gibbs. Elementary Principles in Statistical Mechanics. Charles Scribner's Sons (1902).
- [2] E. T. Jaynes. Information Theory and Statistical Mechanics. Physical Review 106, 620-630 (1957). DOI 10.1103/PhysRev.106.620.
- [3] Claude E. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423 and 623-656 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x; 10.1002/j.1538-7305.1948.tb00917.x.
- [4] Herbert B. Callen. Thermodynamics and an Introduction to Thermostatistics, 2nd ed. Wiley (1985).
- [5] Frank L. Lambert. Disorder - A Cracked Crutch for Supporting Entropy Discussions. Journal of Chemical Education 79(2), 187 (2002). DOI 10.1021/ed079p187.
- [6] Rolf Landauer. Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development 5(3), 183-191 (1961). DOI 10.1147/rd.53.0183.

Causalidad: intervención, contrafactuales e influencia física

IDE-004 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.7 ES · 17 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Ciencia experimental, inferencia causal y propagación física

Límite: La causalidad se separa de correlación, predicción y mero orden temporal

Trabajos relacionados: IDE-005, IDE-010, IDE-023

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Las preguntas causales exigen más que comprobar si dos variables cambian juntas. Preguntan qué cambiaría bajo intervención, qué condiciones de fondo deben controlarse y qué explicaciones alternativas se han excluido. Los experimentos aleatorizados obtienen fuerza causal al modificar deliberadamente la asignación del tratamiento, mientras las ciencias observacionales dependen de supuestos más fuertes sobre estructura, confusión, tiempo y mecanismo.[1,2] La causalidad de Granger formaliza una asimetría predictiva en series temporales, pero predecir no equivale a intervenir.[3] La física añade otra frontera: la relatividad restringe por dónde pueden propagarse influencias en el espacio-tiempo, mientras el teorema de Bell excluye amplias clases de explicaciones locales con variables ocultas sin proporcionar un canal controlable de señalización superlumínica.[4-6] El artículo separa estos significados para que «causa» no se reduzca a correlación, cronología o predictibilidad. La asociación puede sugerir una pregunta causal, pero la conclusión requiere diseño o restricciones que descarten alternativas.

En 60 segundos

Supongamos que fumar y una enfermedad están correlacionados. La correlación no dice por sí sola qué ocurriría si se cambiara el tabaquismo manteniendo adecuadamente fijos los factores de fondo relevantes. Esa pregunta contrafactual o intervencionista es el núcleo de la inferencia causal moderna.[1,2] Los experimentos aleatorizados ayudan rompiendo vínculos sistemáticos entre asignación del tratamiento y muchos confusores potenciales. Cuando aleatorizar es imposible, grafos causales, experimentos naturales, variables instrumentales, información temporal y conocimiento mecanístico pueden a veces identificar efectos, pero solo bajo supuestos explícitos.[1] La prueba de Granger pregunta si el pasado de una serie mejora la predicción de otra más allá de su propio pasado; es útil, pero no demuestra por sí sola un mecanismo manipulable.[3] «A ocurrió antes que B» es aún más débil porque causas comunes, feedback y selección pueden producir secuencias engañosas.[1] En física, la relatividad hace de la velocidad de la luz una velocidad límite, de modo que las influencias controlables no la superan.[4] Las correlaciones de Bell violan desigualdades de modelos locales con variables ocultas, pero la teoría cuántica estándar sigue prohibiendo señalización superlumínica controlable.[5,6] La

causalidad no es, por tanto, un solo estadístico. Es una afirmación disciplinada ligada a intervenciones, estructura del modelo e influencia físicamente permitida.

Por qué importa esta pregunta

El lenguaje causal dirige decisiones: qué tratamiento funciona, qué política cambia el riesgo, qué mecanismo produjo un fallo y si un suceso físico puede influir en otro. Confundir correlación con causa puede recomendar intervenciones inútiles o dañinas; confundir predicción con causa puede hacer que un modelo de pronóstico parezca mecanístico; confundir correlación cuántica con señalización puede convertir un resultado preciso en ciencia ficción. La solución no es exigir experimentos aleatorizados para toda pregunta, sino declarar qué noción causal se usa, qué supuestos la identifican y qué observaciones revelarían confusión o una restricción física violada.

1. La asociación todavía no es un efecto causal

Una asociación estadística describe cómo covarían variables en una distribución observada. Un efecto causal pregunta cómo diferiría un resultado si se cambiara por intervención un tratamiento o entrada frente a una alternativa apropiada.[1,2] Son objetos distintos. Dos variables pueden asociarse con fuerza porque una causa a la otra, porque una tercera causa a ambas, porque la selección depende de ambas o porque el proceso generador contiene feedback.

El lenguaje contrafactual hace explícita la distinción. El marco de resultados potenciales de Rubin define efectos comparando resultados que ocurrirían bajo estados alternativos de tratamiento, aunque para una unidad solo observemos uno.[2] El marco estructural de Pearl expresa intervenciones modificando el modelo causal en vez de limitarse a condicionar un valor observado. [1] Los formalismos difieren en énfasis, pero coinciden en que la distribución conjunta por sí sola no resuelve la pregunta causal.

Incluso condicionar puede crear una asociación inexistente. Si dos causas independientes influyen en si un caso entra en una muestra seleccionada, restringir el análisis a ese grupo puede volverlas estadísticamente dependientes. Este sesgo de collider o selección muestra por qué «controlar por todo» no es una regla causal segura.[1] Las variables de ajuste deben elegirse desde una estructura causal, no desde la intensidad de correlación.

En qué punto estamos

La asociación revela patrones que merecen explicación, pero los efectos causales se refieren a intervenciones o contrastes contrafactuales que la distribución observacional no determina por sí sola. Confusión y selección pueden imitar dirección. El paso siguiente es metodológico: preguntar qué diseños y supuestos proporcionan fuerza suficiente para identificar un efecto causal en vez de redescubrir la misma correlación.

2. Los experimentos ganan fuerza mediante intervención

La asignación aleatoria es potente porque, en expectativa, hace el tratamiento independiente de características previas que podrían confundir la comparación. Las diferencias de resultados pueden atribuirse a la intervención sujetas a límites de adherencia, medición, interferencia y muestreo.[2] La aleatorización no garantiza equilibrio perfecto en toda muestra finita ni responde automáticamente si el resultado generaliza fuera de la población estudiada.

La intervención también separa manipulación de observación. Ver que quienes eligen un tratamiento evolucionan mejor puede reflejar pronóstico, acceso o motivación; asignarlo mediante

protocolo aleatorio cambia el proceso generador de datos. Pearl formaliza la diferencia con operadores de intervención y los resultados potenciales con contrastes entre resultados específicos de tratamiento.[1,2] Ambos obligan a decir qué acción se compara con qué alternativa.

Los experimentos no siempre son posibles o éticos. La astronomía no puede aleatorizar estrellas, la geología no puede repetir colisiones continentales y la epidemiología no puede asignar muchas exposiciones dañinas. La inferencia causal depende entonces de variación cuasialeatoria, experimentos naturales, variables instrumentales, controles negativos, conocimiento mecanístico y supuestos estructurales cuidadosamente defendidos.[1] La falta de aleatorización eleva la carga argumental; no vuelve inaccesible la causalidad.

En qué punto estamos

La intervención proporciona la forma empírica más limpia de una pregunta causal, pero mucha ciencia trabaja sin control experimental directo. En esos casos, la identificación proviene de supuestos que conectan observaciones con una estructura causal. La disciplina crucial consiste en hacer esos supuestos suficientemente visibles para poder atacarlos. Eso abre el problema observacional: distinguir vías causales de confusión mediante grafos, ajustes y diseños alternativos.

3. La causalidad observacional depende de estructura

Los diagramas causales codifican hipótesis sobre qué variables influyen directamente en cuáles. Su valor no es decorativo: criterios gráficos de separación pueden indicar qué variables ajustar, cuáles no condicionar y cuándo un efecto no es identificable desde la distribución observacional disponible.[1] Un grafo combina por tanto información empírica con supuestos que no vienen incluidos en los coeficientes de correlación.

Las causas comunes son los confusores clásicos. Si condiciones socioeconómicas influyen tanto en una exposición como en un resultado, puede persistir una asociación sin efecto causal directo. Pero ajustar tampoco siempre ayuda. Condicionar mediadores puede eliminar parte del efecto que queremos estimar; condicionar colliders puede introducir sesgo.[1] El análisis causal trata del papel de una variable en el proceso generador hipotético, no de si su p-valor es pequeño.

Experimentos naturales y variables instrumentales explotan situaciones donde algún factor desplaza el tratamiento y, bajo supuestos declarados, afecta al resultado solo a través de él. Estos diseños pueden ser potentes, pero sus supuestos de identificación son sustantivos y a menudo no comprobables únicamente con los datos observados.[1] Los análisis de sensibilidad, la replicación entre contextos y la coherencia mecanística ayudan porque obligan a que modos de fallo distintos conspiran para producir el mismo resultado causal aparente.

En qué punto estamos

La inferencia causal observacional es posible cuando la estructura aporta información que la asociación bruta no contiene. Grafos y cuasiexperimentos no eliminan supuestos; los organizan para que ajustes, identificación y fallos sean explícitos. Una fuente importante de estructura adicional es el tiempo, pero el orden temporal tiene su propia trampa: predecir desde el pasado puede ser evidencia útil sin convertirse en causalidad intervencionista.

4. El tiempo ayuda, pero predecir no es intervenir

En intervenciones ordinarias hacia delante, las causas no ocurren después de sus efectos, de modo que la información temporal restringe muchas hipótesis. Pero «A fue primero» no basta. Ciclos

estacionales pueden impulsar A y B, un proceso latente puede producir respuestas retrasadas en ambos y el feedback puede hacer que cada variable informe sobre el futuro de la otra.[1] La precedencia temporal estrecha posibilidades sin seleccionar un mecanismo único.

La causalidad de Granger es deliberadamente predictiva. En su forma clásica, X causa a Y en sentido de Granger cuando el pasado de X mejora la predicción de Y más allá de la información contenida en el pasado de Y, bajo el modelo y conjunto de información elegidos.[3] Puede revelar dependencia direccional y es valiosa en economía, neurociencia y otros datos secuenciales. Pero causas comunes omitidas, retrasos de medición y mala especificación pueden crear o borrar relaciones de Granger.

La distinción es constructiva y no desestimatoria. La asimetría predictiva puede ser una pieza de evidencia causal cuando se integra en un diseño o modelo estructural que trate confusión y medición. Se vuelve engañosa solo cuando la palabra «causalidad» de la etiqueta técnica se toma como prueba de respuesta a una intervención. Un pronóstico puede funcionar por razones que desaparecerían si el predictor se manipulara físicamente.

En qué punto estamos

El tiempo ordena influencias candidatas y la asimetría predictiva puede revelar estructura direccional útil, pero ni cronología ni predictibilidad de Granger identifican por sí solas un efecto de intervención. La causalidad sigue dependiendo del modelo de qué pudo producir las observaciones. La física añade una frontera externa más aguda: la geometría del espacio-tiempo restringe qué sucesos pueden influirse incluso antes de elegir un modelo causal estadístico.

5. La causalidad física tiene una frontera espacio-temporal

La relatividad especial hace de la velocidad de la luz una velocidad límite: componer dos velocidades inferiores a ella da siempre una velocidad inferior a ella, y las velocidades mayores no tienen posibilidad de existencia.[4] Minkowski lo reformuló como geometría. En cada punto del universo un cono de luz separa los sucesos que envían luz hacia él de los que reciben luz de él, y un punto situado entre los dos conos puede considerarse, según el sistema de referencia elegido, anterior, simultáneo o posterior al vértice.[7] Este artículo llama a esa región separada espacialmente, y como esa velocidad límite se cumple,[4] ninguna influencia controlable la atraviesa. Esto va más allá de decir que las causas preceden a efectos: Einstein ya mostró que dos sucesos simultáneos en un sistema de coordenadas no tienen por qué serlo vistos desde otro en movimiento relativo,[4] de modo que el orden causal debe ligarse a la estructura invariante del espacio-tiempo.

El teorema de Bell aporta otra lección. Bell formuló matemáticamente la idea de variables ocultas locales y mostró que es incompatible con las predicciones estadísticas de la mecánica cuántica, por lo que las correlaciones no se explican asignando variables locales preexistentes del modo clásico que él limitó.[5] Su artículo propuso el experimento decisivo en lugar de informar de uno. Pero la no localidad de Bell no es un teléfono operacional superlumínico. La teoría cuántica estándar conserva no señalización: las elecciones en un ala no permiten controlar los resultados marginales observados en otra ala separada espacialmente de ella.[6,7]

Ambas afirmaciones son compatibles porque «causalidad local», dependencia estadística y señalización controlable no son conceptos idénticos. La física refuerza así la regla metodológica general del artículo: definir la afirmación causal antes de probarla. IDE-005 pasa a simetría y

pregunta qué permanece invariante bajo transformación en vez de qué cambia qué. Ese traspaso es editorial y no constituye evidencia para ninguna afirmación de esta sección.

En qué punto estamos

La causalidad abarca contrastes de intervención, identificación estructural en datos observacionales e influencia físicamente permitida en el espacio-tiempo. La disciplina común es excluir alternativas bajo supuestos explícitos. Correlación, precedencia temporal, predictibilidad de Granger y correlación de Bell aportan información, pero ninguna puede rebautizarse sin más como «causa». El concepto conserva su utilidad precisamente porque su significado operacional se fija antes de interpretar la evidencia.

Lo que sabemos y cómo lo sabemos

- Intervenciones aleatorizadas y bien diseñadas pueden identificar efectos causales modificando la asignación del tratamiento en vez de limitarse a observar exposición.[1,2]
- La asociación observada por sí sola no identifica dirección causal porque confusión, selección y feedback pueden producir el mismo patrón estadístico.[1]
- La causalidad de Granger es una relación predictiva en series temporales y no equivale a causalidad intervencionista.[3]
- La relatividad especial hace de la velocidad de la luz una velocidad límite, de modo que la propagación causal controlable no la supera.[4]
- Las correlaciones de Bell excluyen amplias explicaciones locales con variables ocultas mientras la teoría cuántica estándar mantiene no señalización operacional.[5,6]

Lo que todavía no sabemos

- Cómo identificar muchos efectos causales en sistemas biológicos, sociales y ecológicos complejos cuando intervenir es imposible y la confusión latente solo se mide parcialmente.
- Qué abstracciones causales permanecen estables cuando las propias variables cambian de significado entre escalas, contextos o regímenes de medida.
- Cómo integrar mejor modelos mecanísticos, predictores de aprendizaje automático e identificación causal formal sin permitir que el rendimiento predictivo sustituya la validez causal.

Qué podría hacernos cambiar de opinión

- Una intervención aleatorizada que produjera repetidamente ausencia de cambio donde un modelo causal bien especificado predice un efecto grande, con adherencia y medición verificadas, obligaría a revisar el modelo y no la mera correlación observada.[1,2]
- Un método temporal que recuperara de forma única efectos de intervención en sistemas con causas comunes ocultas arbitrarias usando solo orden predictivo derribaría la distinción actual entre predicción de Granger e identificación causal.[1,3]
- Un protocolo reproducible que usara correlaciones de Bell para transmitir un mensaje elegido libremente fuera del cono de luz contradiría la no señalización estándar y exigiría revisar profundamente la física cuántica relativista.[4-7]

Si solo recuerdas una cosa

Una afirmación causal dice qué cambiaría bajo una intervención definida o una influencia físicamente permitida; correlación, cronología y predicción son evidencia, no sustitutos.

Artículos relacionados

- IDE-005 — Simetría
- IDE-010 — Retroalimentación
- IDE-023 — ¿Qué es observación?

Referencias

- [1] Judea Pearl. Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge University Press (2009).
- [2] Donald B. Rubin. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5), 688-701 (1974). DOI 10.1037/h0037350.
- [3] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37, 424-438 (1969). DOI 10.2307/1912791.
- [4] Albert Einstein. On the Electrodynamics of Moving Bodies. *Annalen der Physik* 17, 891-921 (1905). DOI 10.1002/andp.19053221004.
- [5] John S. Bell. On the Einstein Podolsky Rosen Paradox. *Physics Physique Fizika* 1(3), 195-200 (1964). DOI 10.1103/PhysicsPhysiqueFizika.1.195.
- [6] Tim Maudlin. Quantum Non-Locality and Relativity, 3rd ed. Wiley-Blackwell (2011).
- [7] Hermann Minkowski. Raum und Zeit. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 18, 75-88 (1909); address delivered at Cologne on 21 September 1908.

Simetría: transformaciones, invariantes y estados rotos

IDE-005 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.5 ES · 17 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Geometría, dinámica, teoría cuántica y transiciones de fase

Límite: Simetría significa invariancia respecto a una transformación especificada; la regularidad visible es solo un caso especial

Trabajos relacionados: IDE-006, IDE-013, IDE-020

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La simetría afirma qué permanece sin cambio cuando se aplica una transformación especificada. Un copo de nieve sugiere simetría rotacional, pero la física moderna usa una idea más amplia: ecuaciones, acciones, estados o leyes de probabilidad pueden ser invariantes bajo rotaciones, traslaciones, transformaciones internas o intercambios aunque nada parezca geoméricamente regular. Las simetrías continuas pueden restringir la dinámica, y el teorema de Noether conecta simetrías diferenciables de una acción con corrientes conservadas bajo supuestos definidos.[1] La simetría también clasifica posibilidades: Wigner usó representaciones de simetrías espacio-temporales para organizar estados relativistas de partículas.[2] Sin embargo, una ley simétrica no tiene por qué producir un estado realizado simétrico. Ferromagnetos y teorías de campos muestran ruptura espontánea, donde las ecuaciones conservan simetría mientras el estado elegido no.[3-5] La simetría gauge exige otra cautela porque las transformaciones gauge locales representan redundancia descriptiva, de modo que «romper una simetría gauge» no equivale literalmente a romper una simetría global observable.[6]

En 60 segundos

Para llamar simétrico a algo, primero hay que nombrar la transformación y el objeto probado. Un círculo es invariante bajo cualquier rotación sobre su centro; un cristal puede serlo solo bajo ciertas traslaciones y rotaciones discretas; una acción física puede ser invariante aunque una solución individual no lo sea. Las simetrías se componen y su estructura algebraica se describe mediante grupos y representaciones. El primer teorema de Noether muestra que simetrías continuas diferenciables de una acción generan corrientes conservadas cuando se cumplen sus supuestos: invariancia temporal se asocia con conservación de energía, traslaciones espaciales con momento y rotaciones con momento angular.[1] Eso no significa que «el tiempo cause la conservación de energía»; significa que una invariancia concreta de la descripción dinámica tiene una consecuencia matemática precisa. Wigner usó representaciones del grupo de Poincare para clasificar estados de una partícula mediante invariantes como masa y espín.[2] La simetría también puede quedar oculta por el estado. Bajo una temperatura crítica, un ferromagneto puede elegir una dirección de magnetización aunque las interacciones

subyacentes sean rotacionalmente simétricas. En teoría de campos, la ruptura espontánea produce excitaciones características, mientras el mecanismo de Higgs da masa a bosones gauge mediante una estructura más sutil.[3-6]

Por qué importa esta pregunta

La simetría comprime el razonamiento científico. En vez de resolver por separado cada orientación, posición o configuración interna, una invariancia indica qué situaciones son físicamente equivalentes y qué cantidades deben conservarse o clasificarse juntas. Pero esa potencia facilita exagerarla. Una forma bella no basta, una ley de conservación no sigue de cualquier parecido vago y un estado con simetría rota no implica que las ecuaciones hayan perdido su simetría. El lenguaje preciso separa transformación, objeto invariante y estado que la realiza, evitando varias confusiones frecuentes a la vez.

1. La simetría siempre es simetría bajo algo

Una simetría se define respecto a una transformación. Rotar un círculo perfecto en su plano no lo cambia; trasladar un cristal infinito ideal un vector de red conserva su patrón atómico; intercambiar dos partículas idénticas en una descripción cuántica transforma el estado según la simetría de permutación pertinente. Transformación y objeto sobre el que actúa deben especificarse antes de que la palabra simetría tenga contenido científico.

El objeto invariante no tiene por qué ser una forma visible. Una ecuación diferencial puede conservar su forma bajo un cambio de coordenadas; una acción puede ser invariante bajo una familia continua de transformaciones; una distribución de probabilidad puede ser invariante aunque una muestra extraída de ella no lo sea. Así se entiende por qué la física puede ser muy simétrica sin producir imágenes visualmente simétricas.

También conviene separar simetrías exactas, aproximadas y accidentales. Una exacta está incorporada al modelo en el nivel considerado. Una aproximada vale cuando se desprecian efectos menores. Una accidental puede emerger de los operadores disponibles a cierta escala sin imponerse como principio fundamental. Sus consecuencias predictivas difieren, y llamarlas a todas simplemente «simetría» oculta lo que el modelo afirma.

En qué punto estamos

La simetría solo se vuelve precisa cuando se nombran transformación, objeto y nivel de descripción. La regularidad visual es un ejemplo, no la definición. Una vez expresada matemáticamente, la invariancia puede restringir la dinámica y no solo describir apariencia. El caso más célebre es el teorema de Noether, que convierte simetrías continuas de una acción en leyes de conservación bajo supuestos específicos.

2. Noether convierte simetría continua en conservación

El teorema de Noether trata transformaciones continuas diferenciables de un sistema variacional. Si la acción es invariante bajo una transformación de ese tipo, las ecuaciones implican una corriente conservada correspondiente.[1] En mecánica y teoría de campos, invariancia ante desplazamientos temporales se asocia con conservación de energía, traslaciones espaciales con momento e invariancia rotacional con momento angular. Son resultados estructurales, no coincidencias empíricas añadidas tras resolver las ecuaciones.

Los supuestos importan. El teorema se refiere a una acción especificada y su simetría; no dice que todo sistema tenga toda ley de conservación ni que el tiempo fabrique energía. Un accionamiento externo dependiente del tiempo rompe la invariancia temporal de la descripción del subsistema y permite flujo de energía entre subsistema y accionador. Ampliar la frontera puede restaurar conservación para el sistema cerrado combinado.

Noether sirve así como diagnóstico de modelo. Si una acción supuestamente aislada tiene simetría temporal exacta pero sus ecuaciones no producen la cantidad conservada esperada, hay que revisar cálculo, simetría o supuestos variacionales. En sentido inverso, observar una conservación puede motivar buscar una simetría subyacente, pero pasar de una regularidad medida a una simetría fundamental única no es automático.

En qué punto estamos

La simetría continua es más que economía estética: en dinámica variacional puede generar corrientes conservadas. Pero también organiza qué clases de estados son posibles antes de resolver una interacción concreta. La teoría cuántica hace especialmente visible ese papel porque los estados físicos portan representaciones de grupos de simetría. La clasificación de Wigner muestra cómo masa y espín emergen como invariantes del propio espacio-tiempo.

3. La simetría clasifica estados y distinciones permitidas

En teoría cuántica relativista, Wigner estudió representaciones unitarias del grupo de Lorentz inhomogéneo, o grupo de Poincare.[2] La estructura de representación clasifica estados de una partícula mediante etiquetas invariantes. Para partículas masivas, masa y espín aparecen a partir de invariantes de Casimir; para partículas sin masa, helicidad y estructuras relacionadas cumplen el papel correspondiente. La clasificación es cinemática y organiza tipos posibles de estado con independencia de las fuerzas detalladas.

La simetría también restringe transiciones. Muchas reglas de selección en física atómica y nuclear aparecen porque ciertos elementos de matriz se anulan salvo que estado inicial, operador y estado final se combinen de modo permitido por la simetría. La regla puede ser exacta si la simetría lo es o aproximada si campos externos, interacciones débiles o correcciones de orden superior la rompen. Una transición «prohibida» no es metafísicamente imposible; está suprimida respecto a los supuestos usados.

Esta perspectiva invierte una intuición común. La simetría no solo elimina distinciones; dice cuáles son físicamente significativas. Estados que se transforman entre sí pueden pertenecer al mismo multiplete, mientras etiquetas invariantes separan sectores inequivalentes. Clasificación y conservación son dos caras de la misma estrategia: usar invariancia para reducir el número de posibilidades independientes que la teoría debe tratar.

En qué punto estamos

La simetría clasifica estados y puede suprimir procesos además de generar leyes de conservación. Sin embargo, las leyes pueden ser simétricas mientras la naturaleza se instala en uno de varios estados asimétricos. No es una contradicción, sino un cambio en qué porta la simetría: las ecuaciones siguen invariantes mientras el estado realizado selecciona una opción. La ruptura espontánea hace esta distinción experimentalmente concreta.

4. Una ley simétrica puede tener un estado asimétrico

Consideremos un ferromagneto con interacciones microscópicas rotacionalmente simétricas. Por encima de la temperatura de ordenamiento puede no existir dirección preferida. Por debajo, una muestra macroscópica puede elegir una dirección de magnetización, produciendo un estado que ya no es invariante aunque las ecuaciones subyacentes sí. Rotar el estado elegido genera otro estado permitido de igual energía en vez de devolver exactamente el mismo.

En teoría de campos, la ruptura espontánea de una simetría global continua tiene consecuencias características. Goldstone mostró que, bajo supuestos amplios, la simetría continua rota va acompañada de excitaciones sin masa.[3] La lógica importante es que degeneración del vacío y simetría de la teoría restringen juntas el comportamiento de baja energía. La simetría queda oculta en la familia de estados permitidos, no borrada de las ecuaciones.

Esto evita una trampa lingüística. «Simetría rota» suele significar que el estado, vacío o parámetro de orden no comparte toda la simetría de las leyes. La ruptura explícita es distinta: un término adicional de las propias ecuaciones viola la simetría. Los experimentos pueden distinguirlas porque la ruptura explícita suele levantar degeneraciones y modificar el espectro de manera controlada por el término de ruptura.

En qué punto estamos

La ruptura espontánea explica cómo estados asimétricos y nuevos comportamientos colectivos surgen de leyes simétricas. Para simetrías globales, el teorema de Goldstone relaciona esa estructura con modos de baja energía. Las teorías gauge exigen otra precisión: las transformaciones gauge locales son redundancias de descripción, así que el mecanismo de Higgs electrodébil no puede entenderse sin más como una simetría observable ordinaria que elige dirección igual que un ferromagneto.

5. La simetría gauge y el mecanismo de Higgs exigen cuidado extra

Higgs, y Englert con Brout, mostraron en modelos gauge abelianos cómo campos gauge acoplados a un campo con potencial de ruptura adquieren excitaciones masivas de espín uno, y los modos que de otro modo no tendrían masa pasan a ser sus grados de libertad longitudinales.[4,5] En la teoría electrodébil, el campo de Higgs posee un valor esperado del vacío no nulo en una descripción convencional con gauge fijado, y el espectro observado contiene bosones W y Z masivos junto a una excitación física de Higgs. El mecanismo es central, pero su lenguaje se simplifica con facilidad.

Una transformación gauge local cambia variables descriptivas sin cambiar el estado físico, por lo que se comporta de modo distinto a una simetría global ordinaria. El teorema de Elitzur muestra que una simetría gauge local no puede romperse espontáneamente en una formulación gauge-invariante en el sentido ingenuo de un parámetro de orden local.[6] Los físicos siguen usando «ruptura espontánea gauge» como abreviatura eficiente en perturbación con gauge fijado, pero la redundancia gauge sigue siendo redundancia.

La lección más amplia es metodológica. Toda afirmación de simetría debe decir qué se transforma y si la transformación conecta estados físicamente distintos o solo descripciones distintas del mismo estado. IDE-006 pasa de rotaciones y simetrías internas a cambios de escala y coarse-graining. Ese traspaso es conceptual, no evidencia: pregunta qué permanece invariante cuando la transformación es ahora cambio de aumento o resolución.

En qué punto estamos

La simetría es un lenguaje controlado de transformaciones, invariantes, conservación y estructura de estados. Puede ser exacta, aproximada, rota explícitamente u ocultada espontáneamente por el estado

que el sistema selecciona, y la redundancia gauge es una categoría adicional que no debe confundirse con una simetría global. El artículo siguiente formula otra pregunta de invariancia: qué sobrevive a un cambio de escala.

Lo que sabemos y cómo lo sabemos

- Una simetría científica es invariancia de un objeto especificado bajo una transformación especificada, no mera regularidad visual.
- El teorema de Noether conecta simetrías continuas diferenciables de una acción con corrientes conservadas bajo sus supuestos variacionales.[1]
- Las representaciones de la simetría espacio-temporal clasifican estados relativistas de partículas mediante etiquetas invariantes como masa y espín.[2]
- La ruptura espontánea permite que leyes simétricas posean estados realizados asimétricos y, para simetrías globales continuas rotas, modos característicos de baja energía.[3]
- La simetría gauge local es una redundancia descriptiva y requiere lenguaje más cuidadoso que una ruptura espontánea ordinaria; Elitzur impide un parámetro de orden local gauge-invariante ingenuo.[4-6]

Lo que todavía no sabemos

- Qué simetrías aproximadas de física más allá del Modelo Estándar, si existen, son restos de simetrías exactas más profundas a energías hoy inaccesibles.
- Cómo emergen y desaparecen simetrías en fases fuertemente acopladas cuando ningún modelo microscópico débilmente acoplado hace obvios los grados de libertad relevantes.
- Qué principios de simetría son restricciones fundamentales de las leyes físicas y cuáles son regularidades efectivas producidas por dinámicas más profundas o límites de baja energía.

Qué podría hacernos cambiar de opinión

- Un experimento reproducible con un sistema aislado que violara una conservación de Noether mientras la acción probada mantuviera la simetría continua exacta correspondiente y todos los supuestos del teorema se verificaran obligaría a revisar el modelo o la asignación de simetría.[1]
- Una partícula relativista cuyos estados de una partícula no pudieran organizarse consistentemente mediante la estructura de representaciones de la simetría espacio-temporal cuestionaría la clasificación estándar de Wigner usada en teoría cuántica de campos.[2]
- Un observable local gauge-invariante que adquiriera valor esperado no nulo únicamente por ruptura espontánea de una redundancia gauge local exacta, contradiciendo el teorema de Elitzur bajo sus supuestos, exigiría reconsiderar la descripción gauge estándar.[6]

Si solo recuerdas una cosa

Simetría significa invariancia bajo una transformación especificada; conservación, clasificación y estados rotos solo siguen cuando objeto transformado y supuestos están explícitos.

Artículos relacionados

- IDE-006 — Escala
- IDE-013 — Tiempo
- IDE-020 — ¿Qué es una partícula?

Referencias

- [1] Emmy Noether. Invariant Variation Problems. English translation by M. A. Tavel. Transport Theory and Statistical Physics 1, 186-207 (1971). DOI 10.1080/00411457108231446.
- [2] Eugene P. Wigner. On Unitary Representations of the Inhomogeneous Lorentz Group. Annals of Mathematics 40(1), 149-204 (1939). DOI 10.2307/1968551.
- [3] Jeffrey Goldstone. Field Theories with Superconductor Solutions. Il Nuovo Cimento 19, 154-164 (1961). DOI 10.1007/BF02812722.
- [4] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. Physical Review Letters 13(16), 508-509 (1964). DOI 10.1103/PhysRevLett.13.508.
- [5] F. Englert; R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. Physical Review Letters 13, 321-323 (1964). DOI 10.1103/PhysRevLett.13.321.
- [6] S. Elitzur. Impossibility of Spontaneously Breaking Local Symmetries. Physical Review D 12, 3978-3982 (1975). DOI 10.1103/PhysRevD.12.3978.

Escala: ¿qué detalles importan en cada nivel?

IDE-006 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguroola

Versión: v0.7 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Modelización multiescala, coarse-graining y teoría efectiva

Límite: Escala trata la dependencia de variables y leyes útiles respecto a la resolución; emergencia se reserva para IDE-007

Trabajos relacionados: IDE-007, IDE-016, IDE-021

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Las descripciones científicas cambian con la escala porque no todos los grados de libertad importan igual para toda pregunta. Un gas puede describirse mediante trayectorias moleculares, campos hidrodinámicos o variables termodinámicas, y ninguno de esos niveles es simplemente el real para cualquier propósito. La separación de escalas permite resumir detalles más pequeños o rápidos cuando su efecto sobre el fenómeno de interés está controlado, y las teorías efectivas formalizan esa estrategia reteniendo los grados relevantes por debajo de un cutoff elegido.[3] Cerca de transiciones continuas las fluctuaciones abarcan muchas longitudes y la separación simple falla. El grupo de renormalización de Wilson explica cómo el coarse-graining repetido modifica las descripciones efectivas y por qué sistemas microscópicos distintos pueden fluir hacia el mismo comportamiento macroscópico.[1,4,5] Las leyes de escala tienen por tanto dominios, cutoffs y regiones de crossover en vez de validez infinita. Este artículo trata la escala como control de relevancia y deja la organización colectiva genuinamente novedosa para IDE-007.[2]

En 60 segundos

Escala no es solo el tamaño de un objeto. Especifica la resolución, longitud, tiempo, energía o frecuencia característicos de una descripción. Cuando escalas microscópicas y macroscópicas están bien separadas, muchos detalles finos pueden promediarse: los fluidos no necesitan seguir cada molécula, la óptica geométrica no sigue cada oscilación electromagnética y la física de partículas a baja energía puede usar interacciones efectivas sin conocer todo el detalle ultravioleta.[3] La aproximación funciona mientras las escalas omitidas permanezcan suficientemente lejanas y su influencia pueda capturarse sistemáticamente. Cerca de un punto crítico, las longitudes de correlación crecen y aparecen fluctuaciones en muchas escalas a la vez. Las ideas de bloques de Kadanoff y el grupo de renormalización de Wilson muestran cómo el coarse-graining produce flujos en parámetros; los puntos fijos explican universalidad, donde materiales con constituyentes microscópicos distintos comparten exponentes críticos.[1,4,5] Los sistemas reales conservan tamaño finito, espaciado de red, disipación y otros cutoffs, así que leyes de potencia y autosimilitud deben ligarse a un rango medido. «More Is Different» de Anderson subraya algo relacionado: en cada escalón hacen falta leyes, conceptos y

generalizaciones enteramente nuevos, de modo que una ciencia no es sin más otra aplicada.[2] La escala dice cuándo esas variables se estabilizan; la emergencia pregunta qué novedad explicativa poseen.

Por qué importa esta pregunta

Muchas disputas científicas son en realidad desajustes de escala. Un modelo molecular puede ser exacto en principio e inútil para predecir un huracán; un continuo puede describir muy bien el flujo y fallar en un gas rarificado; una ley de potencia puede ajustar tres décadas y romperse en un cutoff físico. Decir «todo se reduce a átomos» no identifica las variables necesarias para la pregunta concreta. El análisis de escala hace explícita la aproximación: qué se resuelve, qué se promedia, cómo entran los errores y dónde la descripción cruza a otro régimen.

1. Las descripciones efectivas tienen regímenes definidos

Una descripción microscópica sigue muchos grados de libertad; una macroscópica conserva menos variables colectivas. En un gas, posiciones y velocidades moleculares pueden sustituirse por densidad, presión, temperatura y velocidad de flujo cuando el equilibrio local y la separación de escalas estabilizan esos campos. Las variables macroscópicas no son resúmenes ficticios: son magnitudes medibles con ecuaciones predictivas, aunque ignoren deliberadamente la mayor parte del detalle microscópico.

La elección de escala sigue la pregunta. La propagación del sonido depende de campos de presión y densidad a distancias mucho mayores que el camino libre medio molecular. La cinética química puede necesitar especies moleculares sin requerir funciones de onda electrónicas para cada átomo. Un modelo climático resuelve celdas de circulación que un modelo meteorológico subdividiría, y ninguno sigue cada colisión del aire. En el análisis del grupo de renormalización, el coarse graining cambia las variables y los acoplamientos útiles y produce descripciones efectivas dependientes de la escala.[5] Más detalle microscópico no equivale automáticamente a más explicación.

El coarse-graining reemplaza distinciones finas por variables agregadas. Puede ser espacial, temporal, energético o estadístico. Un buen coarse-graining preserva los rasgos relevantes para la predicción objetivo y suprime grados de libertad cuyos valores detallados se promedian. El peligro es una dependencia oculta: si variables omitidas retroalimentan con fuerza a las retenidas, el modelo grueso puede fallar aunque sus ecuaciones parezcan suaves.

En qué punto estamos

La escala selecciona qué variables son suficientemente estables para sostener predicción. El coarse-graining gana manejabilidad descartando detalle, pero solo se justifica cuando los grados descartados tienen efectos controlados sobre la descripción retenida. Esta lógica se vuelve sistemática cuando existe una jerarquía de escalas separadas. Las teorías efectivas convierten «ignorar lo pequeño» en una expansión con dominio de validez explícito.

2. La separación de escalas hace posibles las teorías efectivas

Una teoría efectiva describe fenómenos por debajo de una energía característica o por encima de una longitud característica sin resolver cada mecanismo de alta energía. El programa de lagrangianos efectivos de Weinberg organiza términos de baja energía permitidos por las simetrías. [3]

Esto no significa que la teoría sea falsa. La mecánica newtoniana es efectiva respecto a relatividad y cuántica, pero sigue siendo extraordinariamente precisa cuando velocidades son pequeñas frente a

la luz y efectos cuánticos son despreciables. La mecánica de fluidos continua es efectiva respecto a dinámica molecular cuando las longitudes características superan el camino libre medio. A menudo el error puede estimarse como una razón de escalas, y eso convierte la aproximación en ciencia y no solo conveniencia.

La separación puede fallar. Si un grado supuestamente pequeño se vuelve resonante, longevo o fuertemente acoplado a las variables retenidas, integrarlo puede producir grandes errores. Cerca de umbrales, transiciones de fase o crossovers, la propia jerarquía puede cambiar. La respuesta correcta no es insistir en un nivel preferido, sino ampliar la descripción efectiva hasta representar la dinámica relevante.

En qué punto estamos

Las teorías efectivas explotan la separación entre la escala observada y las escalas omitidas. El programa de lagrangianos efectivos de Weinberg organiza términos de baja energía permitidos por las simetrías. Los fenómenos críticos muestran el caso opuesto: fluctuaciones importantes aparecen en muchas escalas y eliminar detalle pequeño ya no deja ecuaciones casi iguales. La renormalización pregunta cómo cambia la propia descripción cuando cambia la resolución.

3. La renormalización sigue cómo cambian las descripciones con la escala

Kadanoff propuso agrupar grados microscópicos en bloques cerca de un punto crítico, generando una nueva descripción efectiva a resolución más gruesa.[4] Wilson convirtió esa intuición en el marco del grupo de renormalización: coarse-graining repetido y reescalado desplazan los parámetros de una teoría por un espacio de descripciones efectivas posibles.[1] El objeto central es un flujo con la escala, no un conjunto único de constantes microscópicas miradas a toda resolución.

Los puntos fijos son especiales porque la descripción macroscópica se vuelve invariante ante nuevo reescalado. Cerca de ellos, pocas direcciones en el espacio de parámetros pueden seguir siendo relevantes mientras muchas diferencias microscópicas se vuelven irrelevantes a larga distancia. [1,5] Esto explica la universalidad: fluidos, imanes y otros sistemas con componentes muy diferentes comparten exponentes críticos si coinciden dimensionalidad, simetría y estructura relevante.

Los términos relevante e irrelevante tienen aquí significado técnico. Un acoplamiento irrelevante no deja de existir; su influencia disminuye bajo coarse-graining cerca del punto fijo. Uno relevante crece y aleja el sistema de criticidad. El vocabulario captura directamente la dependencia de escala y evita tratar la importancia microscópica como propiedad permanente independiente del régimen observado.

En qué punto estamos

La renormalización muestra que cambiar escala puede cambiar qué parámetros importan, y los puntos fijos explican por qué el comportamiento macroscópico olvida muchos detalles microscópicos. La universalidad es así un ejemplo preciso de regularidad robusta entre escalas. Pero la invariancia de escala nunca es automáticamente global: los sistemas de laboratorio tienen tamaño finito, cutoffs microscópicos y parámetros de control que los alejan del punto fijo. Hay que marcar dominios y crossovers.

4. Las leyes de escala necesitan rangos, cutoffs y crossovers

Una ley de potencia como y proporcional a x^a es libre de escala solo en el rango donde el mismo exponente describe realmente los datos. Materiales reales tienen espaciado atómico, tamaño finito, tiempo de observación finito y campos externos. Todo ello introduce límites inferiores y superiores. Un segmento recto en escala log-log puede ser científicamente significativo sin extenderse indefinidamente hacia cero o infinito.

Los crossovers aparecen cuando domina física distinta en escalas distintas. Un polímero puede mostrar un exponente efectivo por debajo de una longitud de persistencia y otro por encima; la turbulencia cambia entre rangos de inyección y disipación; un sistema magnético fuera de la región crítica asintótica puede mostrar exponentes efectivos que derivan hacia valores universales al acercarse a la transición.[5] Publicar un exponente sin su rango puede ocultar el cambio de régimen.

El escalado de tamaño finito convierte una limitación en herramienta. Como un sistema real no desarrolla correlaciones mayores que su tamaño, la dependencia de magnitudes medidas con el tamaño de muestra permite inferir comportamiento crítico del límite infinito.[1,5] La lección metodológica más amplia es que toda afirmación de escala debe incluir intervalo, resolución y condiciones de frontera bajo las que fue establecida.

En qué punto estamos

Las leyes de escala son afirmaciones empíricas y teóricas sobre rangos definidos, no promesas de autosimilitud infinita. Cutoffs y crossovers revelan dónde las escalas omitidas vuelven a influir. La escala se convierte así en disciplina de validez controlada y no en una escalera de «pequeño fundamental» a «grande aproximado». Cuando aparecen variables macroscópicas estables surge otra pregunta: ¿solo resumen la microfísica o sostienen organización explicativa nueva?

5. Escala no es lo mismo que emergencia

Anderson argumentó que reducir todo a leyes fundamentales simples no implica poder partir de esas leyes y reconstruir el universo, porque la hipótesis constructorista se rompe ante las dificultades gemelas de escala y complejidad.[2] Superconductividad, simetría rota y excitaciones colectivas ilustran la idea: conocer las leyes microscópicas es necesario, pero puede no proporcionar directamente las variables que organizan el comportamiento macroscópico. Es una afirmación sobre estructura explicativa, no una violación de la microfísica.

Escala y emergencia se solapan pero no deben fusionarse. La escala dice cómo se transforman las descripciones al cambiar resolución, qué variables pierden relevancia y dónde valen aproximaciones. La emergencia pregunta si patrones colectivos estables, restricciones o leyes efectivas poseen autonomía o novedad explicativa respecto a la descripción inferior. Puede estudiarse dependencia de escala sin adoptar una tesis filosófica fuerte sobre emergencia.

La distinción evita además una jerarquía falsa. Una teoría macroscópica puede ser menos fundamental en ontología microscópica y, a la vez, más precisa, compresiva y controlable para una tarea macroscópica. La hidrodinámica no compite con la mecánica molecular por ser la «teoría verdadera»; responde otra pregunta en un régimen donde las variables colectivas son adecuadas. IDE-007 toma la emergencia desde esa frontera y no define como emergente todo cambio de escala.

En qué punto estamos

La escala gobierna relevancia, aproximación y estabilidad de variables al cambiar resolución. Las teorías efectivas explotan escalas separadas, la renormalización explica flujos y universalidad cuando

muchas escalas interactúan, y las leyes de escala requieren cutoffs y crossovers explícitos. Nada de ello privilegia universalmente un nivel descriptivo. El artículo siguiente formula la pregunta adicional que la escala no responde: cuándo el comportamiento colectivo robusto merece llamarse emergente y no solo coarse-grained.

Lo que sabemos y cómo lo sabemos

- El programa de lagrangianos efectivos de Weinberg organiza términos de baja energía permitidos por las simetrías.[3]
- La renormalización wilsoniana sigue cómo cambian parámetros efectivos bajo coarse-graining y explica puntos fijos y universalidad cerca de fenómenos críticos.[1,4,5]
- Las leyes de potencia e invariancia de escala en sistemas reales valen en rangos finitos con cutoffs y crossovers.[5]
- «More Is Different» de Anderson es un argumento canónico de que la hipótesis construccionista se rompe ante las dificultades gemelas de escala y complejidad, y de que en cada escalón hacen falta leyes y conceptos enteramente nuevos.[2]

Lo que todavía no sabemos

- Cómo elegir automáticamente variables coarse-grained óptimas en sistemas complejos fuera de equilibrio donde no existe separación evidente de tiempos o longitudes.
- Qué leyes de potencia observadas reflejan universalidad genuina de grupo de renormalización y cuáles surgen de mezclas, rangos finitos o selección de datos.
- Hasta dónde puede extenderse el razonamiento de teoría efectiva a sistemas biológicos y cognitivos cuyas variables relevantes cambian mientras el propio sistema se adapta.

Qué podría hacernos cambiar de opinión

- Un sistema con comportamiento crítico supuestamente universal que conservara de forma reproducible una dependencia fuerte de acoplamientos microscópicos que la renormalización predice irrelevantes, tras controlar tamaño finito y crossover, cuestionaría la clase de universalidad asignada.[1,5]
- Una teoría efectiva que fallara sistemáticamente a energías muy inferiores a su cutoff declarado pese a incluir todos los términos permitidos por simetría hasta el orden esperado obligaría a revisar su separación de escalas o power counting.[3]
- Una ley supuestamente libre de escala que se rompiera en el mismo cutoff físico reproducible al mejorar instrumentos y ampliar muestras exigiría sustituir autosimilitud ilimitada por un régimen de escalado acotado.

Si solo recuerdas una cosa

La escala determina qué variables y aproximaciones son útiles a una resolución dada; una buena teoría dice qué promedia, dónde funciona y dónde otra escala toma el control.

Artículos relacionados

- IDE-007 — Emergencia
- IDE-016 — Complejidad
- IDE-021 — ¿Qué es un campo?

Referencias

- [1] Kenneth G. Wilson. Renormalization Group and Critical Phenomena I-II. Physical Review B 4, 3174-3205 (1971). DOI 10.1103/PhysRevB.4.3174; 10.1103/PhysRevB.4.3184.
- [2] P. W. Anderson. More Is Different. Science 177, 393-396 (1972). DOI 10.1126/science.177.4047.393.
- [3] Steven Weinberg. Phenomenological Lagrangians. Physica A 96, 327-340 (1979). DOI 10.1016/0378-4371(79)90223-1.
- [4] Leo P. Kadanoff. Scaling Laws for Ising Models Near T_c . Physics Physique Fizika 2, 263-272 (1966). DOI 10.1103/PhysicsPhysiqueFizika.2.263.
- [5] Nigel Goldenfeld. Lectures on Phase Transitions and the Renormalization Group. CRC Press (1992).

Emergencia: cuando los patrones colectivos se vuelven explicativos

IDE-007 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.8 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Fenómenos colectivos, teorías efectivas y explicación multiescala

Límite: La dependencia de escala se trata en IDE-006; aquí emergencia significa organización colectiva robusta y autonomía explicativa

Trabajos relacionados: IDE-006, IDE-016, IDE-018

Perfil epistémico: Interpretación científica actual

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La emergencia describe situaciones donde variables, patrones o leyes colectivas se convierten en el lenguaje explicativo más eficaz de un sistema con muchos componentes aunque sigan siendo compatibles con la microfísica subyacente. Un vórtice fluido está formado por moléculas, pero no resulta útil seguirlo molécula por molécula; una fase ferromagnética presenta variables colectivas asociadas con la ruptura de simetría.[1,4] La renormalización afina la idea mostrando cómo el comportamiento de gran escala puede volverse insensible a muchos detalles microscópicos y producir universalidad entre materiales distintos.[3] Esa robustez es más que decir que un sistema es complicado. Las descripciones emergentes ganan estatus científico cuando sus variables son medibles, su dinámica reproducible y su dominio de validez enunciabile. No requieren violar leyes de nivel inferior. Este artículo separa emergencia de escala, complejidad e ignorancia y pregunta cómo regularidades colectivas pueden depender de la microfísica y ser parcialmente autónomas en explicación.[1,2,5]

En 60 segundos

Partimos de un sistema de muchos cuerpos. En el nivel microscópico puede haber partículas, espines, células o agentes con reglas locales. A mayor escala aparecen variables colectivas estables: temperatura, presión, magnetización, fase, parámetro de orden, una onda o un modo de red. No son materia adicional; resumen comportamiento coordinado que se vuelve reproducible cuando interactúan muchos grados de libertad. La frase de Anderson «more is different» capta que a niveles superiores pueden ser necesarios conceptos organizadores nuevos.[1] La renormalización muestra una ruta precisa: el coarse-graining repetido puede conducir modelos microscópicos distintos al mismo punto fijo, haciendo que el comportamiento macroscópico universal sea insensible a gran parte del detalle.[3] Goldstone trata la ruptura espontánea de simetría y sus modos colectivos.[4] Emergencia no es sorpresa ni incapacidad para calcular el microestado. Un patrón es científicamente emergente cuando la descripción superior posee variables estables, regularidades predictivas y relación controlada con la dinámica inferior.[2,5] La microfísica limita lo posible, mientras la descripción emergente puede seguir siendo la explicación más corta e informativa a la escala de interés.

Por qué importa esta pregunta

La palabra emergencia se usa a menudo como permiso para dejar de explicar: la consciencia emerge, la vida emerge, el orden emerge. Eso es demasiado débil. La emergencia científica debe añadir restricciones, variables y predicciones, no sustituir mecanismo por etiqueta. En el extremo contrario, insistir en que solo las ecuaciones microscópicas explican algo ignora por qué hidrodinámica, teoría de fases y elasticidad funcionan tan bien. La posición útil pregunta cuándo las descripciones colectivas se vuelven robustas, qué las hace insensibles al detalle microscópico y dónde termina exactamente su autonomía.

1. La emergencia es más que tener muchas partes

Un sistema con muchos componentes no es automáticamente emergente. Una gran pila de granos independientes puede ser complicada de enumerar sin mostrar una nueva ley colectiva. La emergencia resulta interesante cuando las interacciones generan variables macroscópicas o patrones estables que sostienen predicciones que una simple lista de componentes no ofrece. Temperatura, magnetización y velocidad de un fluido son ejemplos porque organizan comportamiento reproducible entre enormes números de grados microscópicos.

La variable superior sigue anclada al sistema inferior. La temperatura se relaciona con propiedades estadísticas del movimiento microscópico; la magnetización con alineamiento colectivo; los campos hidrodinámicos con densidad y momento coarse-grained. Pero el mapa es muchos-a-uno: innumerables microestados corresponden al mismo macroestado. Esa compresión permite ignorar detalle sin volver arbitraria la descripción macroscópica.

Llamar emergente al macroestado afirma por tanto organización y estabilidad explicativa, no novedad sobrenatural. Si cada predicción exigiera reconstruir el microestado exacto, la variable superior aportaría poco. Si muchas configuraciones microscópicas producen la misma dinámica macroscópica, la variable colectiva gana autonomía porque sigue lo que permanece robusto frente a diferencias microscópicas irrelevantes. Anderson formuló el punto general diciendo que en cada escalón resultan necesarias leyes, conceptos y generalizaciones enteramente nuevos, y Laughlin y Pines nombraron la propiedad que lo hace funcionar: la insensibilidad de tales fenómenos a lo microscópico.[1,2]

En qué punto estamos

La emergencia requiere más que multiplicidad o dificultad computacional. La clave es una variable colectiva estable que comprime muchos microestados conservando estructura predictiva. Esto sugiere una prueba de autonomía explicativa: comprobar si realizaciones microscópicas distintas producen la misma ley superior. Renormalización y universalidad ofrecen los ejemplos físicos más claros de esa robustez frente al detalle microscópico.

2. La universalidad vuelve cuantitativa la emergencia

Cerca de transiciones continuas, sistemas con constituyentes microscópicos muy distintos pueden compartir exponentes críticos y funciones de escala. El grupo de renormalización de Wilson lo explica mostrando que el coarse-graining suprime muchos acoplamientos microscópicos mientras un conjunto menor de variables relevantes controla el comportamiento de larga distancia.[3] El punto fijo compartido define una clase de universalidad.

Esto es más fuerte que decir que ignoramos los detalles microscópicos. La teoría predice que muchos de ellos realmente no importan para ciertos observables macroscópicos cerca del punto fijo. Un punto crítico líquido-gas y un modelo magnético pueden mostrar comportamiento matemáticamente relacionado pese a diferencias microscópicas evidentes. La regularidad emergente queda anclada en el flujo de descripciones con la escala.

La universalidad también tiene límites. Dimensionalidad, simetría, alcance de interacción y cantidades conservadas pueden cambiar la clase. Lejos de la región crítica regresan detalles microscópicos y crossovers. Una ley emergente es autónoma solo en un dominio donde los detalles irrelevantes siguen siéndolo. Esta autonomía condicional es más precisa que afirmar que el nivel superior flota libre de sus constituyentes.

En qué punto estamos

La universalidad muestra cómo el comportamiento colectivo puede estar anclado microfísicamente y, a la vez, ser insensible a gran parte de la microfísica. La ley superior no es independiente de todo hecho inferior; depende de rasgos restringidos como simetría y dimensionalidad.

3. La ruptura de simetría crea nuevas variables colectivas

Un ferromagneto muestra cómo una nueva variable se vuelve significativa solo en un estado colectivo. Goldstone trata la ruptura espontánea de simetría y sus modos colectivos. Inferencia: en el marco contemporáneo utilizado por este artículo, un estado colectivo de este tipo se describe mediante un parámetro de orden. Las ecuaciones microscópicas pueden seguir siendo rotacionalmente simétricas mientras el estado realizado selecciona una dirección.[1,4] La variable colectiva organiza la respuesta de todo el material.

Las simetrías continuas rotas también pueden generar excitaciones colectivas cuyas propiedades están limitadas por la estructura de simetría. Los modos de Goldstone son el ejemplo canónico para simetrías globales continuas.[4] Estas excitaciones no son constituyentes fundamentales adicionales del Hamiltoniano microscópico; son patrones de movimiento coordinado que se comportan como grados de libertad efectivos en la teoría de baja energía.

Por eso el lenguaje de partículas puede reaparecer en distintos niveles. Fonones, magnones y otras cuasipartículas se comportan como excitaciones identificables aunque sean estados colectivos de un medio subyacente. Su existencia demuestra que «de qué está hecho el sistema» y «qué entidades describen mejor su comportamiento de baja energía» son preguntas científicas distintas.

En qué punto estamos

Goldstone trata la ruptura espontánea de simetría y sus modos colectivos.[4] Su emergencia sigue limitada por la dinámica microscópica, pero sostiene un vocabulario predictivo nuevo. Esto motiva el caso práctico más fuerte para la explicación superior: cuando la teoría colectiva no solo es más corta, sino más estable y manejable que cualquier descripción microscópica factible del mismo fenómeno.

4. Las leyes superiores pueden tener autonomía explicativa

La hidrodinámica predice ondas, choques y vórtices sin seguir cada molécula. La elasticidad predice tensiones y deformaciones sin resolver la estructura electrónica de cada enlace. Funcionan porque leyes de conservación, equilibrio local y separación de escalas producen ecuaciones cerradas o casi cerradas para variables colectivas. La dinámica superior puede probarse directamente en su propia escala.[2]

Laughlin y Pines describieron ese estado como un protectorado cuántico, cuyas propiedades genéricas de baja energía quedan determinadas por un principio organizador superior y por nada más, e identificaron la propiedad que lo hace funcionar como insensibilidad a lo microscópico.[2] Su conclusión es que para algunos fenómenos las ecuaciones subyacentes son irrelevantes, lo cual dice qué zanján esas ecuaciones y no ordena qué es más fundamental. Los electrones siguen obedeciendo cuántica, pero la fase macroscópica puede tener leyes robustas que no conviene derivar de nuevo desde la función de onda exacta de muchos cuerpos para cada muestra.

La autonomía nunca es absoluta. Cambiar bastante temperatura, dimensionalidad, composición o driving puede hacer fallar la teoría efectiva. Un fluido hidrodinámico se vuelve rarificado; un sólido elástico se fractura; una fase desaparece. La ley emergente es real porque predice de forma reproducible dentro de un régimen y limitada porque ese régimen es físicamente producido, no lógicamente necesario.

En qué punto estamos

Las explicaciones superiores pueden ser autónomas en la práctica porque variables colectivas obedecen ecuaciones estables que apantallan casi todo el detalle microscópico. Esa autonomía es condicional y comprobable, no independencia metafísica. Queda un peligro: llamar emergente a cualquier fenómeno difícil de predecir. Complejidad, caos y dificultad computacional pueden volver dura la predicción sin producir ninguna ley colectiva estable nueva, así que deben mantenerse separados.

5. Emergencia no es sinónimo de complejidad o ignorancia

Un sistema caótico puede ser muy difícil de predecir porque pequeñas diferencias iniciales crecen, pero sus variables relevantes y ecuaciones pueden seguir siendo las mismas en todo momento. Eso es sensibilidad o complejidad, no necesariamente emergencia. A la inversa, una ley macroscópica sencilla como difusión puede ser emergente y fácil de predecir. La dificultad no es necesaria ni suficiente.

La ignorancia tampoco basta. Si no calculamos un sistema de muchos cuerpos por recursos limitados, eso no establece un nuevo nivel de organización. La emergencia necesita estructura positiva. Butterfield formula el requisito como comportamiento novedoso y robusto respecto de alguna clase de comparación;[5] en los casos físicos de este artículo eso aparece como realizabilidad múltiple, insensibilidad al detalle microscópico y un dominio donde el modelo superior cierre suficientemente para hacer predicciones independientes.

La frontera de la serie sigue este criterio. IDE-006 preguntó cómo cambian descripciones con la escala; IDE-016 preguntará cómo partes interactivas generan comportamiento complejo. Emergencia queda entre ambas como afirmación sobre organización colectiva estable. Gana su lugar cuando un nuevo nivel descriptivo es útil por razones que el propio sistema suministra, no porque al investigador se le acabó la capacidad de cálculo.

En qué punto estamos

La emergencia es la aparición de variables y leyes colectivas robustas que dependen de la microfísica sin repetir todo su detalle. Universalidad, ruptura de simetría y teorías efectivas muestran cómo ocurre en ejemplos físicos controlados. El concepto no debe usarse como sinónimo de tamaño, complejidad o ignorancia. Su forma más fuerte es explicativa: una descripción superior se vuelve predictiva porque muchas diferencias microscópicas dejan de importar.

Lo que sabemos y cómo lo sabemos

- En niveles superiores pueden resultar necesarios conceptos organizadores nuevos, y el comportamiento genérico que gobiernan puede ser insensible al detalle microscópico.[1,2]
- La renormalización da cuenta de las leyes de escala mostrando que se siguen cuando las ecuaciones fluyen asintóticamente a un punto fijo, con muchos acoplamientos microscópicos resultando irrelevantes.[3]
- Teorías macroscópicas efectivas pueden ser predictivas y comprobables sin seguir cada grado microscópico.[2]
- La emergencia no exige violar leyes microscópicas; su autonomía está condicionada por régimen y escala.[1,2,5]

Lo que todavía no sabemos

- Cómo identificar sistemáticamente variables emergentes en sistemas fuertemente fuera de equilibrio donde no existe parámetro de orden ni separación clara de escalas.
- Qué formas de organización biológica y cognitiva muestran robustez comparable a universalidad física y cuáles dependen esencialmente de contingencia histórica.
- Hasta dónde puede formalizarse la autonomía explicativa entre disciplinas sin convertir emergencia en una etiqueta filosófica pura o un sinónimo de coarse-graining.

Qué podría hacernos cambiar de opinión

- Una ley supuestamente emergente que variara fuertemente con detalles microscópicos predichos irrelevantes en todas las realizaciones probadas debilitaría la afirmación de realizabilidad múltiple o universalidad.[3]
- Un parámetro de orden propuesto para definir una fase colectiva que no distinguiera fases ni predijera respuesta cerca de la transición entre muestras controladas exigiría otra descripción emergente.[4]
- Una teoría superior que no añadiera ninguna compresión predictiva respecto a simulación microscópica exacta y no poseyera variables estables frente a cambios de microestado debilitaría el caso para llamarla científicamente emergente.[5]

Si solo recuerdas una cosa

La emergencia no es magia ni mera complejidad: es la aparición de variables y leyes colectivas estables que siguen ancladas en la microfísica y se vuelven la mejor explicación a su escala.

Artículos relacionados

- IDE-006 — Escala
- IDE-016 — Complejidad
- IDE-018 — Conexión

Referencias

[1] P. W. Anderson. More Is Different. Science 177, 393-396 (1972). DOI 10.1126/science.177.4047.393.

[2] R. B. Laughlin; David Pines. The Theory of Everything. Proceedings of the National Academy of Sciences 97(1), 28-31 (2000). DOI 10.1073/pnas.97.1.28.

[3] Kenneth G. Wilson. Renormalization Group and Critical Phenomena I-II. Physical Review B 4, 3174-3205 (1971). DOI 10.1103/PhysRevB.4.3174; 10.1103/PhysRevB.4.3184.

[4] Jeffrey Goldstone. Field Theories with Superconductor Solutions. Il Nuovo Cimento 19, 154-164 (1961). DOI 10.1007/BF02812722.

[5] Jeremy Butterfield. Less is Different: Emergence and Reduction Reconciled. Foundations of Physics 41, 1065-1135 (2011). DOI 10.1007/s10701-011-9549-5.

Probabilidad: un cálculo para alternativas inciertas

IDE-008 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.5 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Probabilidad matemática, inferencia estadística y azar físico

Límite: Probabilidad como cálculo e interpretación; medición y observación se tratan en IDE-014 e IDE-023

Trabajos relacionados: IDE-003, IDE-017, IDE-023

Perfil epistémico: Matemática establecida con aplicaciones físicas dependientes de la interpretación

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La probabilidad proporciona un cálculo coherente para razonar sobre alternativas inciertas una vez especificados los sucesos, las variables y los supuestos de fondo. Sus reglas restringen cómo se combinan y actualizan las probabilidades, pero la misma matemática admite interpretaciones distintas: frecuencias a largo plazo, grados de creencia, propensiones o resúmenes de información incompleta. Este artículo separa el núcleo formal de esas interpretaciones. Explica por qué la probabilidad condicional tiene dirección, por qué la regla de Bayes necesita tasas base, cómo difieren las probabilidades de la mecánica estadística de las cuánticas y por qué los resultados de tipo Bell restringen ciertas explicaciones de variables ocultas sin convertir la probabilidad en una sustancia física. La lección central es metodológica: un número de probabilidad nunca se interpreta solo. Su significado depende del espacio de sucesos, los supuestos del modelo y el procedimiento que conecta el número con observaciones y, cuando la probabilidad es condicional, de la información condicionante.

En 60 segundos

Un modelo probabilístico empieza definiendo alternativas y asignando números entre cero y uno que obedecen reglas de consistencia. La probabilidad condicional $P(A|B)$ significa la probabilidad de A dentro de los casos en los que B se toma como dado; en general no es igual a $P(B|A)$. La regla de Bayes conecta ambas direcciones, pero solo mediante probabilidades previas o tasas base. Los métodos frecuentistas enfatizan el comportamiento bajo muestreo repetido, mientras los bayesianos tratan la probabilidad como incertidumbre coherente condicionada a información; ambos usan el mismo cálculo subyacente en muchas aplicaciones ordinarias.[1,2] En mecánica estadística, las probabilidades resumen distribuciones sobre muchas posibilidades microscópicas y conectan predicciones macroscópicas con una especificación incompleta del microestado.[3] En teoría cuántica, las probabilidades entran en las reglas predictivas incluso para estados cuidadosamente preparados, mientras las desigualdades de Bell muestran que las correlaciones que predice la teoría cuántica no pueden reproducirse mediante amplias clases de modelos locales de variables ocultas.[5] La probabilidad organiza así incertidumbre y predicción, pero no nos dice por sí sola si la incertidumbre es epistémica, física o una mezcla de ambas.

Por qué importa esta pregunta

La probabilidad aparece en casi toda ciencia cuantitativa, pero la misma notación puede ocultar preguntas muy distintas. Una prueba médica, una previsión meteorológica, una medida cuántica y un ensemble termodinámico pueden producir números entre cero y uno, aunque el significado evidencial de esos números no sea idéntico. Confundir $P(A|B)$ con $P(B|A)$, ignorar tasas base o tratar una probabilidad de modelo como propiedad intrínseca de un objeto puede producir errores muy seguros de sí mismos. Un lenguaje probabilístico claro no es, por tanto, una trivialidad contable. Forma parte de la lógica mediante la que la evidencia se traduce en inferencia, predicción y decisiones bajo incertidumbre. [1,4]

1. El núcleo formal: sucesos, medidas y consistencia

El núcleo matemático moderno de la probabilidad comienza con un espacio de resultados posibles, una colección de sucesos y una medida de probabilidad que asigna números a esos sucesos. Las restricciones familiares son sencillas pero poderosas: las probabilidades son no negativas, el espacio completo tiene probabilidad uno y las alternativas mutuamente excluyentes se suman. La axiomatización de Kolmogórov hizo explícita esa estructura y separó el cálculo de cualquier interpretación filosófica particular.[2]

Cox llegó al mismo territorio desde otra dirección. Preguntó qué reglas debería obedecer una representación numérica de la plausibilidad si el razonamiento ha de seguir siendo consistente al combinar proposiciones. Bajo supuestos explícitos, el cálculo resultante es equivalente en forma a la teoría ordinaria de la probabilidad.[1] La importancia no es que toda incertidumbre deba ser subjetiva, sino que la consistencia interna restringe con fuerza cómo pueden combinarse proposiciones inciertas.

Una medida de probabilidad se define sobre un espacio de sucesos especificado, y una probabilidad condicional especifica además su condición; la formulación de Cox también condiciona la plausibilidad a proposiciones o conocimiento de fondo.[1,2] Decir que un suceso tiene probabilidad 0,2 no identifica si el número es un parámetro de modelo, una frecuencia límite, un grado posterior de creencia o una predicción cuántica.

En qué punto estamos

La probabilidad tiene un esqueleto matemático estable, pero un valor numérico adquiere significado científico solo cuando se especifican las alternativas y, cuando corresponda, los supuestos condicionantes. Esto separa el cálculo del debate sobre qué representa la probabilidad. El siguiente paso es el condicionamiento, porque muchos errores prácticos nacen no de las reglas de suma sino de cambiar la información disponible sin seguir el cambio de la afirmación probabilística.

2. Condicionamiento, Bayes y la dirección de la evidencia

La probabilidad condicional tiene dirección. $P(A|B)$ pregunta con qué frecuencia o plausibilidad aparece A dentro de los casos en que B se da por supuesto; $P(B|A)$ formula la pregunta inversa. Ambas se conectan mediante la regla de Bayes, $P(A|B) = P(B|A)P(A)/P(B)$, de modo que invertir una condición requiere las tasas base relevantes.[2] Una prueba muy precisa puede, por ello, producir una probabilidad posterior moderada cuando la condición buscada es rara.

No es solo una trampa de aula. La selección experimental, el cribado diagnóstico y el filtrado observacional pueden cambiar la población a la que se aplica una probabilidad. Tversky y

Kahneman documentaron errores sistemáticos en juicios bajo incertidumbre, incluido el descuido de probabilidades previas en algunos contextos.[4] Trabajos posteriores han refinado cuánto dependen esos sesgos de la representación y del diseño de la tarea, pero el punto matemático permanece: la condicional inversa no se obtiene de la directa por sí sola.

Los procedimientos bayesianos y frecuentistas suelen responder preguntas operativas distintas. Un intervalo de confianza se define mediante propiedades de muestreo repetido de un procedimiento, mientras un intervalo creíble bayesiano es una afirmación de probabilidad sobre un parámetro condicionada a un modelo y un prior. En problemas regulares con muchos datos los resultados numéricos pueden converger, pero las interpretaciones siguen siendo distintas. La precisión exige decir qué cantidad se informa, no usar «95 por ciento» como si todos los métodos significaran lo mismo.

En qué punto estamos

El condicionamiento hace que la probabilidad responda a la información, pero también crea uno de sus fallos más comunes: invertir o cambiar silenciosamente la condición. La regla de Bayes muestra exactamente qué cantidades adicionales hacen falta. Aclarada esa estructura inferencial, la siguiente pregunta es física: cuando la probabilidad aparece en mecánica estadística o teoría cuántica, ¿qué parte refleja descripción incompleta y qué parte pertenece al propio marco predictivo?

3. Probabilidad en mecánica estadística

La mecánica estadística usa probabilidad porque los sistemas macroscópicos contienen cantidades enormes de grados de libertad microscópicos. En vez de seguir cada molécula, se trabaja con distribuciones sobre microestados o ensembles y se derivan cantidades macroscópicas mediante promedios. Jaynes subrayó que los métodos de máxima entropía pueden entenderse como inferencia bajo restricciones declaradas, no como la afirmación de que un gas recorra literalmente de forma ingenua todo estado concebible.[3]

La elección del ensemble forma parte, por tanto, del modelo. Las descripciones microcanónica, canónica y gran canónica corresponden a restricciones e intercambios distintos con el entorno. En límites termodinámicos apropiados suelen coincidir para propiedades volumétricas, pero sistemas finitos, interacciones de largo alcance y situaciones fuera del equilibrio pueden mostrar diferencias importantes. Aquí la probabilidad está ligada a coarse-graining, preparación e información macroscópica limitada.

Esto no convierte las regularidades termodinámicas en arbitrarias. Una vez fijados modelo, restricciones y leyes de interacción, las distribuciones de probabilidad producen predicciones precisas sobre fluctuaciones, relaciones de equilibrio y transporte. La distinción es entre una distribución que resume muchos microestados y la afirmación de que la probabilidad sea un ingrediente material dentro del sistema. El éxito estadístico procede de la relación estructurada entre posibilidades microscópicas y observables macroscópicos.

En qué punto estamos

La mecánica estadística muestra cómo la probabilidad puede ser indispensable incluso cuando la microdinámica se trata como determinista: la distribución codifica restricciones, coarse-graining y especificación microscópica incompleta. Ese caso impide identificar sin más probabilidad con aleatoriedad fundamental. La teoría cuántica plantea una frontera más difícil, porque sus

probabilidades permanecen en las reglas predictivas estándar incluso para preparaciones cuidadosamente controladas.

4. Probabilidad cuántica y la frontera de Bell

La teoría cuántica asigna probabilidades a resultados de medida mediante el estado cuántico y la medición realizada. Repetir lo que operacionalmente es la misma preparación produce distribuciones estadísticas estables, aunque en general no se prediga con certeza cada resultado individual. El formalismo coloca así la probabilidad directamente en la conexión entre preparación y resultados observados, no solo en la ignorancia de una trayectoria clásica descontrolada.

El teorema de Bell afina qué puede y qué no puede explicar esas estadísticas. Las desigualdades de Bell se satisfacen en una amplia clase de teorías locales de variables ocultas; la mecánica cuántica predice violaciones de ellas.[5] El resultado no demuestra que «todo sea aleatorio» ni selecciona una única interpretación de la mecánica cuántica. Excluye una conjunción concreta de supuestos usada para derivar las desigualdades.

La distinción importa. Las probabilidades cuánticas pueden interpretarse mediante marcos tipo Copenhague, muchos mundos, bohmianos, colapsos objetivos y otros, que discrepan sobre ontología mientras reproducen gran parte de la misma estructura probabilística de laboratorio. El núcleo restringido experimentalmente es el patrón de resultados y correlaciones. Las afirmaciones sobre qué es en último término la probabilidad pertenecen a una capa interpretativa que no debe presentarse como si un único experimento la hubiera resuelto.

En qué punto estamos

La teoría cuántica introduce la probabilidad en la regla operativa de predicción, mientras las desigualdades de Bell excluyen amplias explicaciones locales de variables ocultas para las correlaciones que predice la teoría cuántica. Ninguno de esos hechos convierte la probabilidad en sustancia ni resuelve toda disputa interpretativa. La frontera es precisa: los experimentos restringen distribuciones y correlaciones; la ontología requiere argumentos adicionales. La sección final vuelve al uso de probabilidades entre disciplinas sin borrar esas diferencias.

5. La probabilidad como herramienta, no como explicación por sí sola

Un modelo probabilístico puede estar perfectamente calculado y ser científicamente pobre si sus variables, clase de referencia o estructura causal son equivocadas. Asignar una probabilidad a lluvia, enfermedad, fallo o resultado de una partícula no explica el mecanismo que produce la distribución. La probabilidad organiza predicciones condicionadas a un modelo; explicar exige estructura adicional sobre dinámica, causas, restricciones o medición.

Por eso importan la calibración y la comprobación del modelo. Si sucesos a los que se asigna probabilidad 0,7 ocurren muy lejos del 70 por ciento en un conjunto adecuado de predicciones comparables, el sistema de pronóstico está mal calibrado aunque no pueda etiquetarse de antemano cada caso individual como «equivocado». Del mismo modo, un posterior puede ser matemáticamente correcto respecto a un prior y una verosimilitud y, aun así, resultar científicamente engañoso si esos ingredientes representan mal el proceso que generó los datos.

Dentro de esta serie, IDE-017 trata la incertidumbre de forma más amplia, incluida la incertidumbre de medida y de modelo, mientras IDE-023 pregunta qué cuenta como observación e IDE-014 cómo la medición produce cantidades comparables. La probabilidad aporta el cálculo que

conecta alternativas y evidencia; no sustituye esos conceptos vecinos. Esa frontera mantiene relacionadas dependencia estadística, incertidumbre, observación y azar físico sin colapsarlas en una sola palabra.

En qué punto estamos

La probabilidad es más fiable cuando el espacio de sucesos, el modelo y la interpretación se declaran juntos, junto con la información condicionante cuando corresponda. El cálculo se comparte entre muchas ciencias, mientras el significado físico de una probabilidad puede variar mucho entre aplicaciones. Eso no es un defecto sino una condición de frontera. La serie puede usar ahora la probabilidad como herramienta matemática controlada sin tratar incertidumbre, causalidad u observación como sinónimos.

Lo que sabemos y cómo lo sabemos

- El cálculo estándar de probabilidad está restringido por axiomas o requisitos de coherencia que determinan cómo se combinan y condicionan probabilidades una vez fijada la estructura de sucesos.[1,2]
- La regla de Bayes requiere tasas base al invertir probabilidades condicionales; $P(A|B)$ no puede inferirse en general de $P(B|A)$ por sí sola.[2]
- La mecánica estadística usa distribuciones sobre microestados o ensembles para derivar predicciones macroscópicas bajo restricciones especificadas.[3]
- Las desigualdades de Bell restringen amplias clases de explicaciones locales de variables ocultas, y la teoría cuántica predice correlaciones que las violan.[5]

Lo que todavía no sabemos

- Qué interpretación de la probabilidad cuántica, si alguna, ofrece la mejor descripción ontológica detrás del formalismo predictivo compartido.
- Cómo representar de forma más eficaz la incertidumbre en sistemas complejos no estacionarios donde el espacio de sucesos relevante y la propia clase de modelos evolucionan.
- Cuándo las discrepancias entre modelos probabilísticos se entienden mejor como falta de datos, mala especificación del modelo o supuestos físicos genuinamente distintos.

Qué podría hacernos cambiar de opinión

- Experimentos cuánticos repetibles y bien calibrados que mostraran desviaciones estadísticamente significativas de las probabilidades de la regla de Born tras excluir sistemáticos conocidos exigirían revisar el marco probabilístico cuántico estándar.
- Un modelo local de variables ocultas que reprodujera las correlaciones que predice la teoría cuántica manteniendo los supuestos usados para derivar la desigualdad derribaría la frontera que aquí se extrae del teorema de Bell.[5]
- Una regla de inferencia demostrada que recuperara $P(A|B)$ de forma única a partir de $P(B|A)$ para distribuciones arbitrarias sin usar priors ni información equivalente de tasa base contradiría la estructura de probabilidad condicional usada aquí.[2]

Si solo recuerdas una cosa

La probabilidad es un cálculo coherente para alternativas inciertas; el número solo significa algo con sus sucesos, condiciones, modelo e interpretación.

Artículos relacionados

- IDE-003 — Entropía
- IDE-017 — Incertidumbre
- IDE-023 — ¿Qué es una observación?

Referencias

[1] R. T. Cox. Probability, Frequency and Reasonable Expectation. American Journal of Physics 14, 1-13 (1946). DOI 10.1119/1.1990764.

[2] A. N. Kolmogorov. Foundations of the Theory of Probability, 2nd English ed. Chelsea Publishing (1956; original German edition 1933).

[3] E. T. Jaynes. Information Theory and Statistical Mechanics. Physical Review 106, 620-630 (1957). DOI 10.1103/PhysRev.106.620.

[4] Amos Tversky; Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. Science 185(4157), 1124-1131 (1974). DOI 10.1126/science.185.4157.1124.

[5] John S. Bell. On the Einstein Podolsky Rosen Paradox. Physics Physique Fizika 1(3), 195-200 (1964). DOI 10.1103/PhysicsPhysiqueFizika.1.195.

Campo: magnitudes distribuidas en un dominio

IDE-009 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.5 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Descripciones continuas desde la escala de laboratorio hasta la astrofísica

Límite: La idea organizadora general de campo; qué es un campo cuántico y por qué las partículas son sus excitaciones corresponde a IDE-021

Trabajos relacionados: IDE-021, IDE-002, IDE-018

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Un campo asigna un valor a cada punto de un dominio: un número, un vector, un tensor o un objeto más elaborado, definido sobre el espacio, sobre el espacio y el tiempo, o sobre otro conjunto de índices. La idea nació como manera de sustituir la acción a distancia por algo local. Las líneas de fuerza de Faraday y las ecuaciones de Maxwell la convirtieron de imagen en teoría dinámica, donde el campo transporta energía y momento y se propaga a velocidad finita.[1,2] La misma construcción organiza hoy la velocidad de un fluido, la temperatura, la tensión y la concentración. Este artículo trata el campo como idea organizadora: qué hace falta para definir uno, qué vuelve poderosa la descripción, cuándo un campo es un constituyente físico y no un recurso contable, y dónde se detiene la idea. Qué es un campo cuántico, y por qué las partículas son sus excitaciones, corresponde a IDE-021.

En 60 segundos

Un campo es una aplicación de un dominio en un conjunto de valores. Nombra el dominio, los valores y las ecuaciones que cumplen, y tienes una teoría de campos. El paso que volvió indispensable la idea fue local: en vez de que un cuerpo tire de otro a través del espacio vacío, cada cuerpo genera un campo y el campo actúa donde está el otro cuerpo.[1,6] Las ecuaciones de Maxwell lo concretaron, y traen una consecuencia que la acción a distancia no puede dar. Las perturbaciones se propagan a velocidad finita, así que entre la emisión y la absorción la energía y el momento tienen que estar en alguna parte, y esa parte es el campo.[2,3] Ese es el criterio que separa un campo con estatuto físico de un simple diagrama útil. Un campo de temperatura o de velocidad es una descripción promediada de grados de libertad moleculares, útil por encima de cierta resolución e indefinida por debajo.[4] Un campo electromagnético no es eliminable del mismo modo: transporta un momento que se mide como presión de radiación.[5] La matemática sola no dice en cuál de los dos casos estás. Lo dicen la dinámica y las leyes de conservación.

Por qué importa esta pregunta

El lenguaje de campos abarca ya tanto terreno que oculta una distinción que conviene conservar. Decir que algo se describe mediante un campo no dice nada sobre si el campo es un constituyente del mundo o

un resumen de otra cosa. Si tratas todo campo como sustancia acabas preguntando de qué está hecho el medio; si tratas todo campo como contabilidad pierdes el contenido físico de la teoría de Maxwell, donde el momento reside realmente en el campo entre una carga y otra. La distinción resuelve además cuestiones más modestas: qué es una condición de contorno, qué puede restringir una ley local y cuánto trabajo está haciendo la palabra local.

1. Qué hace falta para definir un campo

Un campo se especifica con tres cosas: un dominio, un destino y una regla. El dominio es el conjunto de puntos donde la magnitud está definida, normalmente una región del espacio o del espaciotiempo, a veces una superficie, una red o un espacio de configuración. El destino dice qué hay en cada punto: un solo número para un campo escalar como la temperatura o el potencial electrostático, una magnitud dirigida para un campo vectorial como la velocidad de un fluido, un objeto matricial para un campo tensorial como la tensión. La regla es el conjunto de ecuaciones que la asignación debe satisfacer.

El destino no es una elección libre de notación. Fija cómo deben transformarse los valores cuando la descripción se reescribe en otras coordenadas, y esa ley de transformación es lo que convierte al campo en un objeto y no en una tabla de números. Una temperatura no cambia al girar los ejes; las componentes de una velocidad sí, y lo hacen de forma determinada. Escribir bien un campo significa escribir algo cuyo contenido físico sobrevive al cambio de descripción.[3]

La regla suele ser una ecuación diferencial, y esa elección implica un compromiso. Las ecuaciones diferenciales relacionan lo que ocurre en un punto con lo que ocurre en su vecindad inmediata, de modo que en cuanto un campo obedece una de ellas la influencia viaja por el dominio paso a paso en vez de saltar. Las condiciones de contorno e iniciales hacen el resto: la misma ecuación describe una cuerda de guitarra y un modo de cavidad, y la diferencia es el contorno.

En qué punto estamos

Un campo es un dominio, un conjunto de valores con ley de transformación y ecuaciones que conectan puntos vecinos. Ese último elemento no es un tecnicismo. Elegir una regla diferencial es elegir que la influencia se propague por el dominio en vez de saltar, que es justamente la propiedad por la que los físicos prefirieron los campos a las fuerzas a distancia.

2. Localidad: la razón por la que la idea se impuso

La gravitación newtoniana funciona, y no es una teoría de campos en sentido fuerte. Da directamente la fuerza entre dos masas, sin decir nada de lo que hay en medio y sin retardo. Faraday, que desconfiaba de la acción a distancia, describió en cambio los efectos magnéticos mediante líneas de fuerza que llenan el espacio alrededor de un imán, y sostuvo que esas líneas tienen existencia física en el espacio intermedio y no son un diagrama de fuerzas sobre un cuerpo de prueba. Enunció esa conclusión como una opinión que sostenía con vacilación.[1,7]

Maxwell convirtió esa intuición en ecuaciones, y las ecuaciones produjeron algo que la imagen anterior no podía. Admiten soluciones de onda que viajan a una velocidad fijada por dos constantes eléctricas medibles, velocidad que coincidía con la de la luz medida.[2] La luz pasó a ser una perturbación electromagnética, y el campo dejó de ser una manera de hablar de fuerzas para ser aquello cuyo estado se sigue.

La propagación finita obliga a plantear la contabilidad. Si se agita una carga aquí y otra responde después allá, la energía y el momento que cruzaron entre ambas existieron durante el intervalo, y no estaban en ninguna de las dos cargas. La electrodinámica clásica se los atribuye al campo, con una densidad de energía y un flujo de momento definidos, y la atribución no es decorativa: la presión de radiación se mide, y la conservación del momento solo para las cargas falla sin ella.[3,5]

En qué punto estamos

El concepto de campo se impuso porque volvió la influencia local y de velocidad finita, y porque eso obligó a atribuir energía y momento al propio campo. La imagen de Faraday y las ecuaciones de Maxwell pusieron contenido físico en el espacio entre los cuerpos. Pero no todo campo gana ese estatuto, lo que plantea cómo distinguir los dos tipos.

3. Campos efectivos y campos fundamentales

Un campo de velocidad asigna una velocidad a cada punto de un fluido. No existe la velocidad del fluido en un punto matemático: el valor es un promedio sobre un volumen que contiene muchas moléculas pero pequeño frente a la escala del flujo. El campo es un objeto genuino y predictivo dentro de esa ventana y carece de sentido fuera de ella, y por eso las descripciones continuas fallan en gases enrarecidos, donde el recorrido libre medio se vuelve comparable al aparato.[4]

La temperatura se comporta igual. Está definida para una región en equilibrio local, no para una molécula, y un campo de temperatura presupone que el equilibrio se establece localmente más rápido de lo que cambia el propio campo. Cuando esa separación falla, el campo no se vuelve inexacto: se vuelve indefinido, y hay que usar otra descripción.

No se sabe que el campo electromagnético sea el promedio de nada, y no es eliminable de la contabilidad. Es una afirmación sobre la física actual, no una garantía metafísica, y su respaldo es concreto: el campo tiene grados de libertad propios, transporta energía y momento y admite estados, como la radiación que se propaga en el vacío, sin ninguna fuente presente.[3,5] Un campo efectivo puede sustituirse por sus constituyentes; este no.

En qué punto estamos

Unos campos resumen un sustrato y solo están definidos por encima de una resolución; otros son, hasta donde llega la física actual, constituyentes por derecho propio. El criterio no es la matemática, que es común, sino si el campo tiene grados de libertad independientes y magnitudes conservadas. Esa misma matemática explica que la construcción viaje tan lejos del electromagnetismo.

4. La misma construcción fuera de la física

Una vez visible el patrón aparece en todas partes donde una magnitud varía de forma continua sobre un dominio: concentración en una disolución, tensión en una viga cargada, presión en la atmósfera, altitud sobre un terreno, densidad de población sobre un área. En cada caso la disciplina es idéntica. Enuncia el dominio, enuncia qué se asigna, enuncia la ecuación y enuncia la resolución por debajo de la cual la asignación deja de significar algo.

La recompensa es que los métodos de solución se transfieren. La difusión, la conducción del calor y la dispersión de una población en un hábitat uniforme obedecen la misma ecuación con constantes distintas, de modo que un resultado demostrado una vez se reutiliza y una intuición entrenada en uno sirve para otro. Es una economía de pensamiento real y también un riesgo real: compartir

matemática no implica compartir mecanismo, y una ecuación de difusión ajustada a una población en expansión no dice nada sobre por qué se mueven los individuos.

La disciplina también deja al descubierto un error frecuente. Dibujar un campo sobre un dominio donde la magnitud no está realmente definida en cada punto, por ejemplo una superficie suave ajustada a un puñado de estaciones de muestreo, produce una imagen con mucha más estructura de la que sostienen los datos. La interpolación es un modelo, y el campo hereda la incertidumbre del modelo en vez de eliminarla.

En qué punto estamos

La construcción es portátil porque es formal: un dominio, una asignación y una regla local pueden escribirse para casi cualquier magnitud que varíe con continuidad. Esa portabilidad transfiere métodos y transfiere también un riesgo, porque compartir una ecuación no es compartir un mecanismo. Saber qué no resuelve la idea de campo es la última pieza.

5. Dónde se detiene la idea

El concepto de campo no decide por sí solo la ontología. Hesse documenta una transición histórica desde los modelos mecánicos del éter hacia conceptos autónomos de campo electromagnético.[6]

La teoría clásica de campos tiene además dificultades internas que no son solo filosóficas. La energía del campo electromagnético de una carga puntual diverge, y la teoría clásica de una partícula cargada que interactúa con su propio campo no es consistente sin modificaciones.[5] Son problemas del marco clásico, y parte de lo que una teoría más profunda debe abordar.

Esa teoría más profunda es el objeto de IDE-021, que pregunta qué es un campo cuando sus excitaciones están cuantizadas y las partículas aparecen como esas excitaciones. IDE-009 se detiene en la idea clásica y general: la estructura, la razón por la que la localidad la hizo dominante y la diferencia entre un campo que resume y un campo que constituye.

En qué punto estamos

Un campo es una manera disciplinada de poner una magnitud en todo un dominio y dejar que los valores vecinos se restrinjan entre sí. Sustituyó a la acción a distancia, ganó estatuto físico allí donde transporta magnitudes conservadas y sigue siendo un resumen allí donde promedia un sustrato. Qué ocurre cuando ese campo se cuantiza es otra pregunta, y es la siguiente.

Lo que sabemos y cómo lo sabemos

- Un campo se especifica con un dominio, un tipo de valor con ley de transformación definida y ecuaciones que relacionan puntos vecinos; la forma diferencial es lo que hace que la influencia se propague en vez de saltar.[3]
- Las líneas de fuerza de Faraday sustituyeron la acción a distancia por un relato local,[1,7] y la teoría dinámica de Maxwell la convirtió en ecuaciones que dan soluciones de onda que viajan a la velocidad de la luz medida.[2]
- La propagación finita obliga a atribuir energía y momento al campo en tránsito;[3,5] la presión de radiación y el balance de momento hacen esa atribución comprobable y no formal.[5]
- Los campos continuos como la velocidad de un fluido o la temperatura son promedios definidos solo por encima de una resolución fijada por el sustrato, y dejan de estar definidos cuando esa separación de escalas falla.[4]

Lo que todavía no sabemos

- Si las divergencias clásicas asociadas a fuentes puntuales son un artefacto exclusivo del marco clásico o dejan un residuo en las teorías que lo sustituyen.
- Cómo caracterizar mejor la resolución a la que una descripción continua de campo deja de estar definida en sistemas fuertemente forzados o lejos del equilibrio, donde no se establece equilibrio local.
- Si los grados de libertad gravitatorios son en último término de tipo campo en el mismo sentido que los electromagnéticos, cuestión que pertenece al tratamiento cuántico y no al concepto clásico.

Qué podría hacernos cambiar de opinión

- Una medida reproducible de una influencia que se propagase entre cuerpos separados sin energía ni momento transportados por el campo durante el intervalo, descontados todos los portadores conocidos, socavaría la lectura física del campo y devolvería el problema a la acción a distancia.[3,5]
- Una demostración de que la radiación electromagnética en una región sin fuentes puede describirse por completo sin atribuir grados de libertad independientes al campo lo reclasificaría como descripción efectiva, igual que la velocidad de un fluido.[5]
- Experimentos en régimen enrarecido que mostraran que la hidrodinámica continua sigue siendo exacta donde el recorrido libre medio supera la longitud característica invalidarían el relato estándar sobre por qué los campos continuos tienen dominio de validez.[4]

Si solo recuerdas una cosa

Un campo pone un valor en cada punto de un dominio y deja que los vecinos se restrinjan entre sí; que sea constituyente o resumen depende de lo que transporta.

Artículos relacionados

- IDE-021 — ¿Qué es un campo?
- IDE-002 — Información
- IDE-018 — Conexión

Referencias

- [1] Michael Faraday. Experimental researches in electricity — Twenty-eighth series. Philosophical Transactions of the Royal Society of London 142, 25-56 (1852). DOI 10.1098/rstl.1852.0004.
- [2] James Clerk Maxwell. A Dynamical Theory of the Electromagnetic Field. Philosophical Transactions of the Royal Society of London 155, 459-512 (1865). DOI 10.1098/rstl.1865.0008.
- [3] Richard P. Feynman; Robert B. Leighton; Matthew Sands. The Feynman Lectures on Physics, New Millennium Edition, Vol. II, chapters 1-2, 12 and 27 (the field concept; electrostatic analogs; field energy and momentum). Basic Books (2011), ISBN 978-0-465-02416-2.
- [4] G. K. Batchelor. An Introduction to Fluid Dynamics. Cambridge University Press (1967).
- [5] John David Jackson. Classical Electrodynamics, 3rd ed. Wiley (1999), chapters 6, 7 and 16. ISBN 978-0-471-30932-1.
- [6] Mary B. Hesse. Forces and Fields: The Concept of Action at a Distance in the History of Physics. Thomas Nelson and Sons (1961).
- [7] Michael Faraday. On the physical character of the lines of magnetic force. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, series 4, 3(20), 401-428 (1852). DOI 10.1080/14786445208647033.

Retroalimentación: cuando la salida vuelve como entrada

IDE-010 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.6 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Lazos de control en ingeniería, fisiología y sistemas ambientales

Límite: La estructura y la dinámica de los lazos: signo, ganancia, retardo, no linealidad y estabilidad. Cómo se establece un lazo causal a partir de datos observacionales corresponde a IDE-004

Trabajos relacionados: IDE-011, IDE-004, IDE-017

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Hay retroalimentación cuando un proceso modifica una magnitud que después actúa sobre ese mismo proceso. Cerrar el lazo no ajusta simplemente un valor: cambia la dinámica del sistema, y por eso un sistema realimentado puede hacer cosas que ninguna de sus partes hace por separado. Maxwell analizó así el regulador de la máquina de vapor en 1868, y el mismo análisis cubre hoy amplificadores, fisiología, ecosistemas y clima.[1] El artículo expone qué determina el comportamiento de un lazo: el signo del camino de retorno, la ganancia, el retardo y las no linealidades que acaban acotándolo. Sostiene que el signo por sí solo casi no predice nada, que regular no es mantener una consigna y que un diagrama de lazo es una afirmación causal que hay que ganarse, no dibujar.

En 60 segundos

Un lazo se cierra cuando la salida de un proceso vuelve a su propia entrada. La realimentación negativa devuelve una corrección que se opone a la desviación; la positiva devuelve una que la refuerza. Los signos son lo fácil. Lo que decide el comportamiento es la ganancia de lazo, cuánta salida vuelve, y el retardo, cuánto tarda en volver.[3,6] Una realimentación negativa con ganancia alta y suficiente desfase no estabiliza: oscila, y por eso Black tuvo que dar al diseño de amplificadores un criterio de estabilidad y no solo un signo de corrección.[2,3] En modelos de realimentación no lineales, la realimentación positiva acotada puede asentarse en puntos fijos; la biestabilidad y la histéresis dependen de la existencia y estabilidad de múltiples ramas.[5] Regular no es un termostato que sostiene un número: la homeostasis fisiológica mantiene variables dentro de rangos mediante varios lazos que interactúan a velocidades distintas.[4] Y un lazo dibujado en un diagrama afirma dirección y tiempo. Sin eso, un lazo aparente puede disolverse en un impulsor común.

Por qué importa esta pregunta

La retroalimentación es una de las pocas ideas que cruza de verdad la ingeniería, la biología, la economía y la ciencia del clima, y ese alcance la vuelve fácil de usar como adorno. Decir que un sistema

tiene realimentación negativa suena a decir que es estable, y no lo es: eso lo deciden el retardo y la ganancia, y un lazo corrector con los valores equivocados es un oscilador. Decir que tiene realimentación positiva suena a predecir desbocamiento, y normalmente predice un umbral. Acertar aquí importa siempre que lo que está en juego esté en la dinámica y no en la dirección, lo que incluye cualquier discusión sobre con qué rapidez responde un sistema regulado cuando se lo empuja.

1. Qué hace cerrar un lazo

Un regulador centrífugo en una máquina de vapor deja que la propia velocidad de la máquina ajuste el suministro de vapor: si gira rápido, los contrapesos suben y cierran la válvula. La aportación de Maxwell no fue el dispositivo, que ya se usaba, sino el análisis. Escribió las ecuaciones de la máquina junto con su regulador y estudió las raíces que determinan si una perturbación decae o crece.[1]

Ese paso es toda la materia en miniatura. Antes de cerrar el lazo, la velocidad de la máquina es consecuencia del suministro de vapor. Después, ambas quedan mutuamente determinadas, y lo que hay que resolver ya no es un valor sino el comportamiento de un sistema cuyas ecuaciones contienen ahora su propia salida. Los lazos cambian por tanto el orden y el carácter de la dinámica, y añaden modos que no tienen ni la planta ni el controlador por separado.[6]

Por eso la realimentación compra propiedades que el diseño en lazo abierto no puede. Un amplificador cuya ganancia depende de la temperatura y de la válvula concreta montada puede hacerse de modo que su ganancia quede fijada casi por completo por un divisor pasivo en el camino de retorno, porque el lazo lleva el error hacia cero en vez de confiar en el camino directo. [2] La precisión se compra con estructura y no con componentes.

En qué punto estamos

Cerrar un lazo convierte la salida del sistema en parte de su propia causa, lo que cambia las ecuaciones y no solo los números que producen. Eso permite que la realimentación dé exactitud con piezas inexactas. Pero introduce también una dinámica que antes no existía, y si esa dinámica se asienta o crece depende de magnitudes que el signo del lazo no contiene.

2. El signo no basta: ganancia, retardo y fase

La realimentación negativa se opone a la desviación respecto del punto de operación. De ahí no se sigue que la amortigüe. La corrección llega con retardo, y para cuando llega el sistema ya se ha movido. Si el retardo es suficiente para que la corrección empuje en la misma dirección que la perturbación cuando aterriza, el lazo refuerza aquello que se construyó para oponerse.[3,6]

Bode dio la manera clásica de enunciarlo. Sigue la respuesta del lazo frecuencia a frecuencia; allí donde la señal vuelve invertida por el desfase acumulado, pregunta si la ganancia de lazo es mayor o menor que la unidad. Si supera la unidad en esa frecuencia el sistema no se asentará, y los márgenes con que un diseño evita esa condición son las medidas estándar de cuánto retardo o cuánta variación de ganancia tolera.[3,6]

El enunciado de ingeniería generaliza, pero no a todo lazo. La inestabilidad de lazo cerrado y la oscilación sostenida dependen de las condiciones de ganancia y fase.[3,6]

En qué punto estamos

Un lazo negativo estabiliza solo para ciertas combinaciones de ganancia y retardo, y fuera de ellas el mismo lazo se convierte en oscilador. La inestabilidad de lazo cerrado y la oscilación sostenida

dependen de las condiciones de ganancia y fase. La realimentación positiva se presenta a menudo como el caso opuesto, pero su comportamiento habitual tampoco es lo contrario de la estabilidad.

3. Realimentación positiva, umbrales y saturación

La realimentación positiva devuelve una corrección que se suma a la desviación, de modo que las diferencias pequeñas crecen. Un crecimiento a ritmo proporcional al valor actual es exponencial, y si nada interviniera la variable divergiría. En modelos de realimentación no lineales, la realimentación positiva acotada puede asentarse en puntos fijos; la biestabilidad y la histéresis dependen de la existencia y estabilidad de múltiples ramas.[5] Eso es lo que vuelve útil la realimentación positiva en una célula que debe comprometerse con una decisión.

En qué punto estamos

La biestabilidad y la histéresis dependen de la existencia y estabilidad de múltiples ramas.[5] Esta distinción también aclara qué mantienen estable los sistemas regulados.

4. Regular no es mantener una consigna

Cannon acuñó homeostasis para el mantenimiento fisiológico de las condiciones internas y fue cuidadoso con la palabra. Eligió un término que significa condición semejante y no condición fija, porque lo que la fisiología mantiene es un rango dentro del cual el organismo funciona, sostenido por varios lazos con velocidades y efectores distintos.[4]

Un termostato doméstico es un mal modelo de esto. Compara una temperatura medida con un número y conecta un calefactor, y el número se impone desde fuera del lazo. La temperatura corporal se defiende con sudoración, tiritona, cambios vasomotores y conducta, cada uno con su propio umbral, y el rango defendido se desplaza con la hora del día, la actividad y la enfermedad. No hay un único objetivo almacenado que se esté siguiendo.[4]

La distinción importa más allá de la fisiología. Describir un sistema regulado como poseedor de una consigna invita a preguntar dónde está guardado el número, y para casi toda regulación natural no hay respuesta porque no hay número. Lo que existe es una estructura de lazos cuya acción conjunta tiene una región estable, y esa región es propiedad del sistema entero y no un valor contenido en una de sus partes.

En qué punto estamos

Regular suele significar una región estable mantenida por varios lazos, no un objetivo almacenado que uno solo persigue. La homeostasis se nombró por la condición semejante y no por la fija, y el propio rango defendido se mueve. Queda entonces la cuestión de cómo se establece que un lazo dado existe realmente, que es donde los diagramas de realimentación deben ganarse.

5. Un diagrama de lazo es una afirmación causal

Dibujar una flecha de A a B y otra de B a A afirma dos influencias dirigidas y un orden temporal. Ninguna de las dos se obtiene de observar que A y B se mueven juntas. Dos variables impulsadas por una tercera común se correlacionan, y su correlación puede parecer un lazo al representarla sin que ninguna actúe sobre la otra.

Establecer un lazo exige por tanto la misma evidencia que cualquier afirmación causal: intervención donde es posible, o supuestos estructurales explícitos donde no lo es, más una temporalidad que distinga los caminos. La ingeniería de control suele tener la versión fácil, porque

el lazo puede abrirse deliberadamente y estudiarse el sistema con el retorno desconectado. Las disciplinas que no pueden abrir el lazo tienen que trabajar mucho más, y ese problema corresponde a IDE-004.

La versión honesta de una afirmación sobre realimentación nombra por tanto la variable del lazo, el signo, la ganancia, el tiempo de respuesta y la cota del crecimiento, y dice cómo se estableció la dirección. Un diagrama con flechas y sin números es una hipótesis dibujada con forma sugerente.

En qué punto estamos

Un diagrama de lazo afirma dirección y tiempo, y ninguna de las dos se sigue de la correlación, así que una afirmación sobre realimentación hay que ganarla con intervención o con supuestos estructurales explícitos. Nombrar la variable, el signo, la ganancia, el retardo y la cota de saturación es lo que convierte el diagrama en modelo. IDE-011 recoge un proceso que cierra un lazo propio, donde la persistencia diferencial altera lo que viene después.

Lo que sabemos y cómo lo sabemos

- Cerrar un lazo cambia la dinámica del sistema y no solo sus valores de salida, y por eso la realimentación puede dar una exactitud que no depende de la precisión del camino directo.[1,2]
- Que un lazo negativo estabilice depende de la ganancia y del retardo, y el criterio en el dominio de la frecuencia para la frontera es práctica estándar de diseño, no una aproximación.[3,6]
- La regulación fisiológica mantiene un rango mediante varios lazos con velocidades y efectores distintos, y el propio rango defendido varía, así que no es una consigna almacenada que se persiga.[4]

Lo que todavía no sabemos

- Cómo identificar de forma fiable la estructura de lazos en sistemas que no pueden abrirse experimentalmente y donde los lazos candidatos operan en escalas de tiempo parecidas.
- Cuánto de la robustez aparente de las grandes redes reguladoras biológicas y ecológicas procede de la arquitectura de lazos y cuánto de la redundancia de los propios componentes.
- Qué medidas cuantitativas de intensidad de realimentación se transfieren de forma útil entre sistemas de ingeniería, donde la ganancia de lazo se diseña, y sistemas naturales, donde hay que inferirla.

Qué podría hacernos cambiar de opinión

- Un sistema de control físico que mantuviera la estabilidad con ganancia de lazo superior a la unidad en la frecuencia donde el desfase acumulado invierte la señal devuelta, contabilizados todos los retardos y dinámicas, invalidaría el criterio de estabilidad estándar.[3,6]
- Una demostración de que una variable fisiológica defendida se sigue frente a un único objetivo numérico almacenado, identificado como magnitud física en una estructura concreta y no inferido del comportamiento, sustituiría el relato de la homeostasis por lazos múltiples.[4]
- Un sistema bien caracterizado con realimentación positiva cuya relación entrada-salida de estado estacionario medida cortara la condición de equilibrio en más de un valor estable, de modo que el modelo ajustado predijera un umbral y un ciclo de histéresis de anchura enunciada, y que aun así mostrara un único estado estable sin ninguno de los dos en barridos lentos de ida y vuelta de su parámetro impulsor, rompería el relato estándar de por qué los lazos positivos dan umbrales.[5]

Si solo recuerdas una cosa

La retroalimentación convierte la salida de un sistema en parte de su propia causa; el signo casi no dice nada, y lo dicen la ganancia, el retardo y qué detiene el crecimiento.

Artículos relacionados

- IDE-011 — Selección
- IDE-004 — Causalidad
- IDE-017 — Incertidumbre

Referencias

- [1] James Clerk Maxwell. On Governors. Proceedings of the Royal Society of London 16, 270-283 (1868). DOI 10.1098/rspl.1867.0055.
- [2] Harold S. Black. Stabilized Feedback Amplifiers. Bell System Technical Journal 13(1), 1-18 (1934). DOI 10.1002/j.1538-7305.1934.tb00652.x.
- [3] Hendrik W. Bode. Network Analysis and Feedback Amplifier Design. Van Nostrand (1945).
- [4] Walter B. Cannon. Organization for Physiological Homeostasis. Physiological Reviews 9(3), 399-431 (1929). DOI 10.1152/physrev.1929.9.3.399.
- [5] Steven H. Strogatz. Nonlinear Dynamics and Chaos, 2nd ed. CRC Press (2015), chapters 3 and 8.
- [6] Karl Johan Åström; Richard M. Murray. Feedback Systems: An Introduction for Scientists and Engineers, 2nd ed. Princeton University Press (2021).

Selección: cómo la persistencia diferencial acumula estructura

IDE-011 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.7 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Poblaciones de variantes que se replican, biológicas y de otro tipo

Límite: La selección como proceso: sus condiciones, qué es la eficacia y qué más cambia las frecuencias. La dinámica de lazo entre una población y el entorno que la filtra corresponde a IDE-010

Trabajos relacionados: IDE-012, IDE-007, IDE-010

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Hay selección cuando variantes de algo difieren en su capacidad de persistir o reproducirse en un entorno dado y esa diferencia se transmite. Tres condiciones hacen el trabajo: variación, éxito diferencial y herencia de aquello que marca la diferencia.[2] Donde se cumplen las tres, la estructura puede acumularse sin previsión y sin diseñador, porque cada generación parte de una población ya desplazada por la anterior. Este artículo enuncia con precisión las condiciones, sostiene que la eficacia es una relación entre una variante y un contexto y no una propiedad que la variante lleva consigo, y expone qué más mueve las frecuencias: la deriva, que es muestreo, y la mutación, que aporta variación sin estar dirigida por la necesidad.[4,5] Cierra con lo que hay que nombrar antes de exportar el lenguaje de la selección fuera de la biología.

En 60 segundos

La selección exige tres cosas y no una. Tiene que haber variación entre entidades, esas variantes deben diferir en cuántos descendientes o copias aportan, y la diferencia debe heredarla lo que viene después; quita cualquiera de las tres y el proceso se detiene.[2] Price mostró que el cambio resultante se escribe exactamente como una covarianza entre un carácter y la eficacia relativa más un término de transmisión, lo que es una identidad contable y no un mecanismo: se cumple sea cual sea la biología, y dice dónde mirar, no qué ocurrió.[3] La eficacia misma es relacional. La eficacia relativa se define respecto de la media de una población.[6] La selección tampoco es lo único que cambia las frecuencias. Las poblaciones finitas derivan por puro muestreo, y buena parte de la variación molecular parece moldeada así.[5] Y la variación llega antes que el filtro: las mutaciones de resistencia bacteriana están en el cultivo antes de añadir el fago.[4]

Por qué importa esta pregunta

La selección es la idea más exportada de la biología, y casi todos los problemas vienen de usar la palabra sin las condiciones. Afirmar que algo fue seleccionado es afirmar que existieron variantes, que

difirieron en éxito y que la diferencia se transmitió, y cada una de esas cosas puede comprobarse o faltar. Saltárselas convierte un mecanismo explicativo en una manera de decir que lo que existe está ahí por alguna razón. La misma disciplina importa cuando el lenguaje viaja a la cultura, los mercados o los algoritmos, donde las tres condiciones a veces se cumplen y a veces solo se insinúan.

1. Las tres condiciones

Lewontin formuló los requisitos en la forma que se ha mantenido: una población presenta variación entre sus miembros, las variantes difieren en eficacia y la eficacia es heredable.[2] Los enuncia para individuos y fenotipos y después los generaliza de forma explícita, señalando que no se especifica ningún mecanismo concreto de herencia y que los principios se aplican por igual a genes, organismos, poblaciones y especies, y por eso el mismo esquema puede aplicarse en principio a otros sistemas que se replican.[2]

Cada condición es una afirmación empírica distinta. La variación puede faltar porque un carácter esté fijado en la población. El éxito diferencial puede faltar porque las variantes rindan igual en el entorno que de hecho se da. La heredabilidad puede faltar porque la diferencia sea adquirida: un organismo que crece más con una dieta rica no produce por ello descendencia mayor, y la población no se desplaza.

El argumento de Darwin trata la variación heredable, la lucha y la supervivencia diferencial, y la selección natural.[1]

En qué punto estamos

La selección necesita variación, éxito diferencial y heredabilidad, y cada una es una afirmación distinta que puede fallar. El esquema no fija ningún nivel, y Lewontin lo aplica por igual a genes, organismos, poblaciones y especies, y eso es lo que le permite viajar. Cumplidas las tres condiciones, la pregunta siguiente es qué magnitud cambia realmente y cuánto, lo que exige precisar la palabra eficacia.

2. La eficacia es una relación, no una propiedad

Price dedujo una expresión exacta del cambio del valor medio de un carácter entre generaciones: es la covarianza entre el carácter y la eficacia relativa, más un término que recoge con qué fidelidad se transmite el carácter.[3] Conviene enunciar el resultado con cuidado porque suele sobreinterpretarse. Es una identidad, cierta por construcción para cualquier población que pueda partirse así, de modo que no es prueba de que hubo selección y no contiene biología propia.

Lo que sí aporta es una descomposición. Si la media se desplazó, o covarió con la eficacia o la transmisión estuvo sesgada, y ambas cosas pueden examinarse por separado. Es una ganancia real, y es también la razón de que la ecuación reaparezca en las discusiones sobre niveles de selección, donde la misma covarianza puede calcularse a nivel de individuos y de grupos.

El carácter relacional de la eficacia sale del mismo sitio. La eficacia relativa se define respecto de la media de una población.[3,6]

En qué punto estamos

La ecuación de Price da una descomposición exacta del cambio en una covarianza con la eficacia y un término de transmisión, y su condición de identidad es lo que la vuelve herramienta y no prueba. Queda abierto cuánto del cambio observado se debe realmente a la selección.

3. La selección no es lo único que ocurre

Las frecuencias cambian en poblaciones finitas incluso cuando las variantes tienen el mismo éxito, porque cada generación es una muestra finita de la anterior. Eso es la deriva, el muestreo aleatorio de variantes en una población finita, y el tamaño poblacional efectivo determina el ritmo de cambio en la composición que la deriva causa: cuanto menor es, mayor la fluctuación y más probable que una variante se pierda o se fije por razones ajenas a sus efectos.[7]

Kimura sostuvo que buena parte del cambio a nivel molecular puede resultar de la fijación aleatoria, por deriva genética, de mutaciones neutras o casi neutras.[5] La tesis se discutió al plantearse y su alcance sigue debatiéndose, pero el punto estructural es ya estándar: observar que una variante se volvió común no es por sí mismo prueba de que fuera ventajosa.

Por eso la selección es una hipótesis y no una opción por defecto. Demostrarla exige mostrar el éxito diferencial y no solo el resultado, y entre las alternativas que hay que excluir están la deriva, el ligamiento con algo distinto que sí fue seleccionado y la posibilidad de que el carácter sea subproducto de una restricción y no una diana.

En qué punto estamos

La deriva cambia las frecuencias por puro muestreo, y su intensidad crece al disminuir la población, de modo que la prevalencia no prueba ventaja. El argumento de Kimura convirtió eso en supuesto de trabajo a nivel molecular y no en curiosidad. Si la selección filtra variación, queda por saber de dónde sale esa variación y si el filtro contribuye a producirla.

4. La variación llega antes y no la dirige la necesidad

Sería económico que los organismos produjeran las variantes que necesitan. Luria y Delbrück diseñaron un experimento que decide la cuestión. Si la resistencia a un fago surge solo como respuesta a la exposición, el número de colonias resistentes entre cultivos réplica debería seguir una distribución estadística; si los mutantes resistentes aparecen espontáneamente antes de la exposición, las mutaciones tempranas producen clones grandes y la varianza entre cultivos debería ser mucho mayor.[4]

Las distribuciones observadas fueron del segundo tipo. Las mutaciones de resistencia estaban en los cultivos antes de introducir el fago, y el fago seleccionó entre ellas en vez de inducirlas. Merece la pena detenerse en el experimento porque convierte una pregunta de aire filosófico en una diferencia medible entre dos distribuciones.

La forma general del resultado es que variación y selección son pasos separados con causas separadas. La mutación aporta candidatos sin atender a lo que sería útil; el entorno filtra después lo que ya estaba. Eso es lo que hace de la selección una explicación del diseño aparente y no una repetición de él, porque la apariencia de propósito la produce un filtro que actúa sobre material no dirigido.

En qué punto estamos

Luria y Delbrück convirtieron la pregunta de si la variación está dirigida en una diferencia entre dos distribuciones, y la respuesta fue que los mutantes resistentes precedían al fago. Variación y selección son por tanto pasos separados con causas separadas, que es justo lo que permite a un filtro producir diseño aparente. Si ese esquema se transfiere fuera de la biología es la última cuestión.

5. Exportar la idea

El esquema es en principio neutral respecto al sustrato, y las analogías con la cultura, los mercados, los repertorios inmunitarios y los algoritmos de optimización son a veces exactas y a veces decorativas. La manera de distinguirlas es nombrar las tres condiciones en el caso concreto: qué varía, qué determina el éxito diferencial y qué transporta la diferencia hacia adelante.

Algunos casos pasan sin problema. Un repertorio inmunitario genera diversidad y amplifica después los linajes que se unen al antígeno, con la diferencia heredada por la descendencia de la célula amplificada. Un algoritmo genético incorpora las tres condiciones por construcción. Otros no: una práctica cultural puede difundirse sin mecanismo alguno que transmita la propiedad responsable de esa difusión, y entonces falla la tercera condición y lo que ocurre es difusión, no selección.

Reaparecen dos modos de fallo. Uno es tomar la persistencia como prueba de eficacia, lo que da por supuesto lo que había que mostrar. El otro es dejar el entorno sin especificar, lo que vuelve la afirmación incontrastable porque la eficacia no tiene valor hasta nombrar el escenario. Ambos se evitan con la misma disciplina: enunciar la población, la variación, el filtro y el canal de herencia.

En qué punto estamos

El lenguaje de la selección se transfiere solo donde las tres condiciones pueden nombrarse en el caso concreto, y los fallos habituales son tomar la supervivencia por prueba de ventaja y dejar el entorno sin enunciar. Nombrar qué varía, qué filtra y qué transporta la diferencia es toda la disciplina. IDE-012 recoge otra pregunta sobre estructura transportada: cómo puede un estado representar algo distinto de sí mismo.

Lo que sabemos y cómo lo sabemos

- La selección exige variación, éxito diferencial y heredabilidad de la diferencia; el esquema no fija ningún nivel y se aplica por igual a genes, organismos, poblaciones y especies, y por eso puede aplicarse a otros sistemas que se replican.[1,2]
- La ecuación de Price descompone el cambio de un carácter medio en una covarianza con la eficacia relativa más un término de transmisión; es una identidad, así que organiza una explicación en vez de proporcionarla.[3]
- La eficacia relativa se define respecto de la media de una población.[3,6]
- La variación se aporta antes del filtro y con independencia de él: los mutantes bacterianos resistentes están presentes antes de aplicar el agente selectivo, como muestra el test de fluctuación.[4]

Lo que todavía no sabemos

- Qué fracción de la variación molecular observada es efectivamente neutra, cuestión que sigue discutida y que varía con el organismo, la región genómica y el tamaño poblacional.
- Con qué frecuencia y en qué condiciones la selección que actúa por encima del nivel del individuo produce una diferencia medible, frente a poder redescibirse en un nivel inferior.
- Qué analogías culturales y tecnológicas satisfacen la condición de heredabilidad de un modo que sostenga predicción cuantitativa y no solo un relato cualitativo.

Qué podría hacernos cambiar de opinión

- Una repetición del test de fluctuación con distribuciones de recuento de colonias compatibles con mutaciones surgidas solo tras la exposición, en condiciones que excluyan el retardo fenotípico y el crecimiento residual previo, invalidaría la afirmación de que la variación precede al filtro.[4]
- Un caso demostrado en el que la media de un carácter cambiara entre generaciones con covarianza nula entre carácter y eficacia relativa y sin sesgo de transmisión, correctamente medidos, contradiría la descomposición de Price e indicaría un error en la partición de la población.[3]
- Evidencia sistemática de que las tasas de fijación en sitios sin efecto detectable sobre la eficacia se apartan de lo que predice la deriva para el tamaño efectivo de población obligaría a revisar el relato neutro del cambio molecular.[5]

Si solo recuerdas una cosa

La selección es variación más éxito diferencial más herencia de la diferencia; sin nombrar las tres, decir que algo fue seleccionado no explica nada.

Artículos relacionados

- IDE-012 — Representación
- IDE-007 — Emergencia
- IDE-010 — Retroalimentación

Referencias

- [1] Charles Darwin. On the Origin of Species by Means of Natural Selection. John Murray, London (1859).
- [2] Richard C. Lewontin. The Units of Selection. Annual Review of Ecology and Systematics 1, 1-18 (1970). DOI 10.1146/annurev.es.01.110170.000245.
- [3] George R. Price. Selection and Covariance. Nature 227, 520-521 (1970). DOI 10.1038/227520a0.
- [4] Salvador E. Luria; Max Delbrück. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28(6), 491-511 (1943).
- [5] Motoo Kimura. Evolutionary Rate at the Molecular Level. Nature 217, 624-626 (1968). DOI 10.1038/217624a0.
- [6] John Maynard Smith. Evolution and the Theory of Games. Cambridge University Press (1982).
- [7] Brian Charlesworth. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10(3), 195-205 (2009). DOI 10.1038/nrg2526.

Representación: cuando una cosa está por otra

IDE-012 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.9 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Mapas, medidas, símbolos, modelos y códigos neuronales

Límite: La propia relación de sustitución. Cuánta información transporta una señal corresponde a IDE-002; si un modelo es adecuado para una inferencia dada corresponde a IDE-015

Trabajos relacionados: IDE-013, IDE-002, IDE-015

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Una representación es algo que se usa para estar en lugar de otra cosa. Hacen falta cuatro elementos antes de que la palabra sirva de algo: un vehículo, lo que sustituye; un objetivo, lo sustituido; una correspondencia entre distinciones de uno y distinciones del otro; y una tarea a la que esa correspondencia debe servir. Representar no exige parecerse, y por eso una fotografía, una ecuación diferencial y un patrón de actividad neuronal pueden representar el mismo objeto.[1,4] Este artículo sostiene que las representaciones son selectivas por construcción y no por defecto, que la estructura estadística no es contenido semántico, y que el lenguaje representacional se gana su sitio solo cuando la correspondencia puede especificarse en vez de suponerse. Si un modelo es adecuado para una inferencia dada se deja a IDE-015.

En 60 segundos

Nada representa nada por sí solo. Una curva en una página se vuelve representación solo cuando alguien la usa con un propósito,[4] y por eso este artículo trata la representación como una relación de cuatro lugares, añadiendo la correspondencia, y no como una propiedad del vehículo. Aceptado eso, varias confusiones se disuelven. El parecido desaparece: la utilidad de un mapa procede de una correspondencia sistemática entre sus distinciones y las del terreno, no de parecerse al suelo, y toda proyección distorsiona algo para conservar otra cosa.[6] La selectividad deja de ser un defecto: una representación que lo conservara todo sería la cosa misma, así que lo que importa es si lo omitido era irrelevante para la tarea. Y la cantidad no es contenido: la teoría de Shannon mide cuánta incertidumbre resuelve un mensaje y dejó de lado los aspectos semánticos como irrelevantes para el problema de ingeniería,[3] de modo que un recuento alto de bits no dice nada sobre de qué trata una señal. Por último, la representación es una hipótesis sobre un sistema, no una descripción libre: atribuirle a un mecanismo es comprometerse con una correspondencia que puede ser falsa.

Por qué importa esta pregunta

La representación es donde varias disciplinas se encuentran y hablan sin entenderse. La ciencia cognitiva pregunta si un estado cerebral representa un borde; la metrología, qué sustituye una lectura; la cartografía, qué conserva una proyección; el aprendizaje automático, qué codifica un vector aprendido. Las cuatro preguntan por la misma relación y las cuatro se equivocan de las mismas dos maneras: suponen una correspondencia sin enunciarla y tratan la pérdida de detalle como imperfección y no como el objetivo. Explicitar la relación es lo que permite decir cuándo una representación es errónea y no solo distinta.

1. Cuatro elementos, no uno

El análisis mínimo tiene cuatro partes. Hay un vehículo: marcas sobre papel, una tensión, una estructura de datos, un patrón de disparo. Hay un objetivo: el terreno, la temperatura, el objeto en el campo visual. Hay una correspondencia que conecta distinciones disponibles en el vehículo con distinciones del objetivo. Y hay una tarea, porque una correspondencia adecuada para un propósito puede ser inútil para otro.[4]

La formulación de Giere hace explícitos al usuario y al propósito en vez de tratar la representación como relación de dos lugares entre una cosa y lo que retrata.[4] Eso importa porque la versión de dos lugares invita a preguntar qué hace que una marca trate intrínsecamente de algo, pregunta sin buena respuesta. La versión de cuatro lugares la sustituye por preguntas que pueden zanjarse: qué distinciones se corresponden con cuáles y si esa correspondencia sostiene la inferencia que se hace.

El parecido queda entonces visiblemente al margen. Un esquema de cableado no se parece a un circuito; una ecuación diferencial no se parece a un cuerpo que cae; la palabra gato no se parece a un gato. Lo que estos casos comparten no es semejanza[4] sino una correspondencia sistemática adecuada a la tarea.

En qué punto estamos

La representación necesita vehículo, objetivo, correspondencia y tarea, y enunciar los cuatro convierte una pregunta irresoluble sobre aboutness intrínseca en preguntas resolubles sobre correspondencia y adecuación. El parecido no es necesario ni suficiente. De ahí se sigue de inmediato que toda representación deja algo fuera, y que eso es una decisión de diseño y no una carencia.

2. Selectiva por construcción

Una representación que conservara todas las distinciones de su objetivo sería un duplicado y no resultaría más manejable que el original. La selectividad es por tanto constitutiva: hay que dejar caer algo para que la representación sirva de algo, y la pregunta interesante nunca es si se perdió detalle sino si el detalle perdido importaba para la tarea.

Las proyecciones cartográficas lo concretan porque la distorsión es inevitable y no solo práctica. Representar un mundo tridimensional en un mapa plano lo distorsiona necesariamente. Elegir una proyección es, por tanto, elegir para qué servirá el mapa. El punto de Monmonier es que eso convierte cada mapa en un argumento y que los argumentos suelen ser invisibles para quien da la proyección por supuesta.[6]

La misma estructura aparece siempre que se elige una representación. Un estadístico resumen deja caer la distribución; un esquema deja caer la geometría; una etiqueta categórica deja caer el gradiente. Cada una es defendible para ciertas inferencias y engañosa para otras, y la manera de llevar la cuenta es nombrar la tarea antes de juzgar la representación.

En qué punto estamos

Toda representación omite, porque una que no omitiera nada sería una copia, así que la cuestión es si la omisión es inofensiva para la tarea. Las proyecciones muestran directamente el compromiso: representar un mundo tridimensional en un mapa plano lo distorsiona necesariamente. Una confusión emparentada no atañe a lo omitido sino a lo conservado, y en concreto a si conservar estructura equivale a conservar significado.

3. La estructura no es contenido

Shannon abrió su artículo de 1948 dejando de lado las cuestiones semánticas, con el argumento de que el problema de ingeniería de reproducir un mensaje en otro punto es independiente de que el mensaje signifique algo.[3] Fue una restricción deliberada del alcance y tuvo un éxito extraordinario, pero implica que las magnitudes resultantes no miden contenido.

Las consecuencias son fáciles de enunciar y fáciles de olvidar. Una cadena aleatoria puede tener entropía máxima y no representar nada. Un mensaje corto puede ser muy significativo para un receptor que posea la convención adecuada y no significar nada para quien no la tenga. Dos archivos pueden tener estadísticas idénticas y tratar de cosas distintas. Nada de esto es una limitación de la teoría de Shannon: es el límite que la teoría se trazó a sí misma.

Para la representación, la moraleja es que una relación estadística entre un vehículo y un objetivo no es suficiente. En la interpretación funcional utilizada aquí, las representaciones cognitivas forman parte de la organización funcional o computacional del sistema y no son meras correlaciones asignadas por un observador externo.[5] Una representación también puede estar causada por algo que no representa, y puede representar algo que no la ha causado.[7]

En qué punto estamos

Las medidas de Shannon se construyeron para funcionar sin semántica, así que la estructura estadística de un vehículo no lo vuelve por sí sola algo que trate de nada. Queda entonces en qué nivel de descripción debe identificarse una representación.

4. Niveles, y por qué la implementación no fija el contenido

Marr sostuvo que un sistema de procesamiento de información debe entenderse en tres niveles: qué problema se resuelve y por qué, qué representación y algoritmo se emplean, y cómo se realizan físicamente.[1] Los niveles están relacionados pero no se reducen unos a otros, y un relato completo del nivel inferior no entrega el superior.

Esa asimetría tiene una consecuencia directa. Conocer todas las tensiones de un circuito o todos los potenciales de acción de una red dice qué hace el sistema físicamente sin decir qué está computando, porque el mismo proceso físico puede implementar cálculos distintos bajo correspondencias distintas. El contenido queda fijado en el nivel donde se especifica la correspondencia, no en el nivel donde corre el mecanismo.

Por eso también el mismo cálculo puede ejecutarse en hardware radicalmente distinto, y por eso las discusiones sobre si un sistema representa realmente algo no se zanján mirando más de cerca el sustrato. Se zanján, si acaso, mostrando de qué correspondencia depende el propio funcionamiento del sistema.

En qué punto estamos

Los niveles de Marr separan el problema, la representación y el algoritmo, y la realización física, y un relato completo del mecanismo no fija qué se computa. El contenido vive donde se especifica la correspondencia. Queda una cuestión práctica: cuándo merece la pena postular una representación en vez de describir el mecanismo directamente.

5. Cuándo se gana su sitio el lenguaje representacional

Newell y Simon propusieron que un sistema físico de símbolos, un sistema con estructuras que designan y procesos que operan sobre ellas, posee los medios necesarios y suficientes para la acción inteligente general, y lo ofrecieron como hipótesis empírica y no como definición.[2] Plantearlo como empírico es lo útil: vuelve el vocabulario representacional algo que puede acertar o errar sobre un sistema dado.

Las descripciones representacionales y computacionales adquieren contenido empírico cuando explican la organización cognitiva más allá de una descripción puramente física.[5]

Ignorar el criterio produce dos fallos. La sobreatribución lee representaciones en toda correlación, lo que vacía la noción. La infraatribución las niega siempre que pueda describirse un mecanismo, lo que pierde la generalización de que mecanismos distintos pueden implementar la misma correspondencia. Nombrar vehículo, objetivo, correspondencia y tarea evita ambos, y es la misma disciplina con la que abría el artículo.

En qué punto estamos

Las descripciones representacionales y computacionales adquieren contenido empírico cuando explican la organización cognitiva más allá de una descripción puramente física. Sobreatribución e infraatribución se evitan nombrando los cuatro elementos. IDE-013 recoge un objetivo que las representaciones organizan constantemente y examinan poco: el tiempo.

Lo que sabemos y cómo lo sabemos

- Este artículo analiza la representación como una relación entre vehículo, objetivo, correspondencia y tarea; explicitar al usuario y el propósito es el movimiento de Giere, y sustituye preguntas irresolubles sobre aboutness intrínseca por otras resolubles.[4]
- La selectividad es constitutiva y no un defecto; las proyecciones cartográficas hacen inevitable el compromiso, porque representar un mundo tridimensional en un mapa plano lo distorsiona necesariamente.[6]
- Shannon dejó de lado los aspectos semánticos de los mensajes como irrelevantes para el problema de ingeniería,[3] así que la estructura estadística de un vehículo no establece por sí sola de qué trata.
- Un relato completo de la implementación física no determina qué se computa, porque el mismo proceso puede realizar cálculos distintos bajo correspondencias distintas.

Lo que todavía no sabemos

- Qué criterios distinguen de forma fiable una representación interna genuina de una correlación que el sistema no usa, en sistemas biológicos donde el uso posterior es difícil de rastrear.
- Cómo caracterizar el contenido de las representaciones distribuidas aprendidas por sistemas artificiales, donde la correspondencia no se diseña y puede no descomponerse en partes interpretables.

- Si el vocabulario representacional es imprescindible para explicar alguna capacidad biológica o es siempre un resumen cómodo de un mecanismo que podría describirse directamente.

Qué podría hacernos cambiar de opinión

- Una demostración de que dos sistemas físicos con dinámica interna idéntica y entradas y salidas idénticas deben recibir el mismo contenido computacional colapsaría la distinción de niveles de Marr y haría que la implementación bastara para fijar el contenido.[1]
- Un criterio operativo que identificara el contenido representacional solo a partir de la estadística del vehículo, sin referencia al uso posterior ni a la tarea, y que sobreviviera a casos donde la misma estadística cumple funciones distintas, invalidaría el análisis de cuatro lugares. [3,4]
- Un sistema biológico bien caracterizado cuyo comportamiento dependiera de un sustituto de un objeto ausente sin correspondencia identificable entre distinciones del vehículo y del objetivo mostraría que el análisis es incompleto y no solo difícil de aplicar.[5]

Si solo recuerdas una cosa

Una representación es un vehículo usado, bajo una correspondencia, para una tarea; sin nombrar esos elementos, llamar representación a algo no describe nada.

Artículos relacionados

- IDE-013 — Tiempo
- IDE-002 — Información
- IDE-015 — Modelo

Referencias

- [1] David Marr. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W. H. Freeman (1982), chapter 1.
- [2] Allen Newell; Herbert A. Simon. Computer Science as Empirical Inquiry: Symbols and Search. Communications of the ACM 19(3), 113-126 (1976). DOI 10.1145/360018.360022.
- [3] Claude E. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423 and 623-656 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x; 10.1002/j.1538-7305.1948.tb00917.x.
- [4] Ronald N. Giere. How Models Are Used to Represent Reality. Philosophy of Science 71(5), 742-752 (2004). DOI 10.1086/425063.
- [5] Zenon W. Pylyshyn. Computation and Cognition: Toward a Foundation for Cognitive Science. MIT Press (1984).
- [6] Mark Monmonier. How to Lie with Maps, 3rd ed. University of Chicago Press (2018).
- [7] David Pitt. Mental Representation. The Stanford Encyclopedia of Philosophy, Spring 2020 edition, Edward N. Zalta (ed.), first published 30 March 2000, substantive revision 21 January 2020. <https://plato.stanford.edu/entries/mental-representation/>

Tiempo: relojes, orden y la flecha

IDE-013 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguro

Versión: v0.10 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Del reloj de laboratorio a los intervalos cosmológicos

Límite: El tiempo como concepto físico y operativo: cómo se define, por qué depende del camino y de dónde viene su dirección. La entropía en sí corresponde a IDE-003; el presupuesto de incertidumbre de un reloj como instrumento corresponde a IDE-014

Trabajos relacionados: IDE-014, IDE-003, IDE-023

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La física no define el tiempo como un flujo. Define una unidad fijando la frecuencia de una transición atómica especificada, y todo lo demás es comparación contra eso.[1] De esa base operativa se siguen tres resultados que la intuición ordinaria no aporta. La simultaneidad de sucesos separados en el espacio depende del estado de movimiento de quien la juzga, así que no hay un ahora universal.[2] El tiempo transcurrido es propiedad de un camino por el espaciotiempo y no del intervalo entre dos sucesos, y relojes en caminos distintos registran totales realmente distintos, algo que los relojes ópticos de laboratorio resuelven hoy sobre una diferencia de altura de decenas de centímetros.[3,5] Y la dirección del tiempo no está en las ecuaciones microscópicas, casi enteramente reversibles: entra por las condiciones de contorno.[6,7] El artículo separa tiempo coordinado, tiempo propio y dirección termodinámica, que suelen confundirse.

En 60 segundos

El segundo del SI se define fijando la frecuencia de una transición hiperfina del cesio-133,[1] lo que convierte la duración en una comparación reproducible y no en la apelación a un reloj universal invisible. Una vez que el tiempo es lo que miden los relojes, la relatividad tiene consecuencias que no son opcionales. Que dos sucesos separados ocurran a la vez depende del movimiento del observador, de modo que la simultaneidad es relativa a un sistema y no un hecho sobre los sucesos.[2] El tiempo propio se acumula a lo largo de una línea de universo, así que dos relojes que se separan y se reencuentran pueden marcar totales distintos sin que ninguno falle, y la gravedad también cambia los ritmos: un reloj más hondo en un potencial gravitatorio va más lento, algo que Pound y Rebka detectaron en una torre y que los relojes ópticos muestran hoy sobre un tercio de metro.[3,4,5] La flecha es otra cuestión. Las leyes microscópicas son esencialmente simétricas ante inversión temporal, así que la asimetría observada de registros, mezcla y memoria tiene que venir de otro sitio, y la respuesta estándar la sitúa en las condiciones de contorno del universo y no en la dinámica.[6,7]

Por qué importa esta pregunta

El tiempo coordinado y el tiempo propio son distintos en relatividad; las flechas termodinámica y psicológica o epistémica son también nociones temporales distintas. El tiempo coordinado es una

etiqueta en una descripción elegida; el tiempo propio es lo que acumula un reloj concreto; la dirección termodinámica es una asimetría de procesos; y el tiempo psicológico es otra cosa más. La dilatación temporal parece una ilusión hasta que se separa tiempo propio de tiempo coordinado; la flecha parece misteriosa hasta que se separa de la dinámica; y la navegación por satélite sencillamente falla si se ignora la distinción.

1. El tiempo es lo que miden los relojes

La definición moderna del segundo no describe qué es el tiempo. Fija un número: la frecuencia de la transición hiperfina del estado fundamental no perturbado del cesio-133 se define como exactamente 9 192 631 770 hercios, con lo que el segundo queda como consecuencia derivada de esa estipulación.[1] La definición es operativa por diseño, y su virtud es que cualquier laboratorio competente puede realizarla.

Es una retirada deliberada de la metafísica, y es el mismo movimiento hecho con el metro cuando se fijó por definición la velocidad de la luz. Nada en el SI afirma que el tiempo fluya, que el presente se desplace o que el futuro no exista. Lo que proporciona es un patrón reproducible de comparación, y la física trata después de relaciones entre esas comparaciones.

La retirada tiene consecuencias. Como la duración se define mediante un proceso físico, cualquier afirmación sobre el tiempo puede contrastarse construyendo relojes y comparándolos. Preguntas que de otro modo serían incontrastables, como si el tiempo transcurrido depende de la ruta seguida, se vuelven preguntas experimentales con respuestas numéricas y barras de error.[5]

En qué punto estamos

La física define el segundo fijando una frecuencia atómica, así que el tiempo es lo que miden los relojes y la duración es una comparación reproducible y no una descripción de flujo. Esa elección operativa es lo que vuelve contrastables las preguntas difíciles. La primera de ellas es si dos sucesos separados pueden siquiera decirse simultáneos.

2. No hay un ahora universal

El argumento de Einstein de 1905 parte de un procedimiento y no de una filosofía. Para decir que dos sucesos separados son simultáneos hace falta una regla que sincronice relojes distantes, y la regla natural usa señales luminosas. Adoptarla, y exigir que la velocidad de la luz sea la misma en todo sistema inercial, fuerza la conclusión de que observadores en movimiento relativo discrepan sobre qué sucesos separados son simultáneos.[2]

La discrepancia no es un error de medida ni una limitación de los instrumentos. El juicio de cada observador es correcto dentro de su propio sistema, y no hay experimento que identifique uno de ellos como el verdadero. La simultaneidad de sucesos separados es por tanto relativa al sistema, mientras que el orden de sucesos conectables por una señal posible no lo es, lo que preserva el orden causal.

La consecuencia para el tiempo en conjunto es que no puede definirse un presente universal. Las secuencias a lo largo de una sola línea de universo siguen perfectamente ordenadas, así que nada de la experiencia personal se vuelve paradójico. Lo que desaparece es el supuesto de que ese orden se extiende hacia fuera, a todo lo demás, de manera única.

En qué punto estamos

La simultaneidad de sucesos separados depende del sistema, así que no puede definirse un presente universal, mientras que el orden de los sucesos conectables causalmente sobrevive intacto. Nada de la experiencia a lo largo de una línea de universo queda amenazado. De ahí se sigue que el tiempo transcurrido debe adscribirse a caminos y no a pares de sucesos, y eso resulta medible.

3. El tiempo transcurrido pertenece a un camino

En relatividad la magnitud que acumula un reloj es el tiempo propio a lo largo de su propia línea de universo, calculado a partir del intervalo espaciotemporal.[3] Dos relojes que parten juntos, siguen rutas distintas y se reencuentran pueden por tanto marcar totales distintos sin que ninguno haya fallado. La asimetría es geométrica: las rutas tenían longitudes distintas en el sentido pertinente.

La gravedad entra por la misma vía. En relatividad general, un reloj más bajo en un potencial gravitatorio acumula tiempo propio más despacio que otro más alto. Pound y Rebka detectaron el desplazamiento de frecuencia correspondiente a lo largo de la altura de una torre de laboratorio en 1960, aprovechando la agudeza de una resonancia nuclear para resolver un desplazamiento fraccional del orden de unas pocas partes en diez a la quince.[4]

Los relojes ópticos han vuelto el efecto casi doméstico. Chou y colaboradores compararon dos relojes de ion de aluminio y resolvieron la diferencia gravitatoria de ritmo producida al elevar uno de ellos unos treinta y tres centímetros, y midieron además el efecto de velocidades relativas de menos de diez metros por segundo.[5] Sea lo que sea el tiempo, su ritmo depende de forma medible de dónde está un reloj y de cómo se mueve, y los sistemas de navegación por satélite deben corregir ambas cosas para funcionar.

En qué punto estamos

Los relojes acumulan tiempo propio a lo largo de su propio camino, así que rutas distintas dan totales distintos, y el potencial gravitatorio cambia el ritmo. Pound y Rebka lo midieron en una torre y los relojes ópticos lo resuelven hoy sobre un tercio de metro. Eso zanja ritmo y camino, y deja intacta la pregunta de por qué el tiempo tiene dirección.

4. La flecha no está en la dinámica

Las ecuaciones fundamentales de la mecánica, el electromagnetismo y la teoría cuántica son simétricas ante la inversión de la coordenada temporal, con una excepción. Las interacciones débiles violan la conjugación de carga y la paridad del modo más fuerte posible y, aunque la combinación CP se conserva en la mayoría de los procesos débiles, se viola en algunos que implican interferencia; se ha observado violación de la inversión temporal, y es corolario de esa violación de CP dado que CPT es una simetría de la naturaleza.[7] Proyecta hacia atrás una película de moléculas que chocan y cada fotograma sigue obedeciendo las leyes. Y sin embargo los registros se acumulan en una dirección, no se observa que lo mezclado se desmezcle, y recordamos el pasado.

La resolución estándar no añade una ley nueva. Observa que la asimetría que vemos es una asimetría de condiciones iniciales: el universo empezó en un estado de entropía muy baja, y el comportamiento posterior de los sistemas grandes se sigue de esa condición de contorno junto con una dinámica simétrica.[6] La dirección del tiempo, según este relato, se hereda de un hecho sobre cómo empezaron las cosas y no de las ecuaciones.

La aportación de Price es insistir en la simetría del razonamiento: una explicación que funcionaría igual de bien aplicada a la dirección futura no ha explicado nada, y buena parte del razonamiento

informal sobre la flecha introduce de contrabando un sesgo temporal desde el principio.[6] El relato es estándar, pero su estatuto es interpretativo y no cerrado, y por qué se dio ese estado inicial de baja entropía sigue siendo una pregunta abierta y no un resultado.

En qué punto estamos

La dinámica microscópica es esencialmente simétrica en el tiempo, así que la flecha observada se atribuye a una condición de contorno de baja entropía y no a una ley. Ese relato es estándar pero interpretativo, y por qué el estado inicial fue así sigue abierto. Lo que sí está claro es que se han llamado tiempo varias cosas distintas, y la última tarea es mantenerlas separadas.

5. De qué tiempo se está hablando

El tiempo coordinado es una etiqueta que una descripción elegida adscribe a los sucesos. Es un recurso contable, y elecciones distintas son igualmente válidas, por lo que un enunciado en tiempo coordinado no tiene contenido físico hasta nombrar el sistema de coordenadas. El tiempo propio es lo que acumula un reloj concreto en un camino concreto, y es invariante: todos coinciden en cuánto registró ese reloj.[3]

La dirección termodinámica no es ninguna de las dos. Es una propiedad de los procesos y no del eje temporal, y es lo que autoriza la distinción entre antes y después en un sistema macroscópico.[6] El tiempo psicológico es otra cosa más, con memoria, atención y construcción de la experiencia, y no es en absoluto una magnitud física.

Inferencia: algunas paradojas temporales aparentes surgen al confundir estas nociones. El enigma de los gemelos mezcla tiempo coordinado con tiempo propio. La afirmación de que la relatividad abole el presente mezcla la simultaneidad dependiente del sistema con el orden a lo largo de una línea de universo. La sospecha de que la física no tiene sitio para el fluir del tiempo mezcla los sentidos termodinámico y psicológico con el geométrico.

En qué punto estamos

IDE-014 recoge qué es un reloj como instrumento, donde la lectura debe convertirse en un valor con su incertidumbre.

Lo que sabemos y cómo lo sabemos

- El segundo del SI se define fijando la frecuencia de la transición hiperfina del cesio-133.[1] Que eso haga de la duración una comparación reproducible y no una descripción de flujo es la lectura que hace este artículo de la definición, y no una afirmación del folleto.
- La simultaneidad de sucesos separados en el espacio depende del sistema de referencia, mientras que el orden de los sucesos conectables causalmente no, de modo que la estructura causal sobrevive a la pérdida del presente universal.[2]
- El tiempo transcurrido es tiempo propio a lo largo de una línea de universo y depende del movimiento y del potencial gravitatorio; el efecto se ha medido en una torre de laboratorio y sobre una altura de unos treinta y tres centímetros con relojes ópticos.[3,4,5]
- La dinámica microscópica es simétrica ante inversión temporal salvo en los procesos débiles que violan CP, así que la flecha termodinámica observada se atribuye de forma estándar a una condición de contorno de baja entropía y no a las ecuaciones.[6,7]

Lo que todavía no sabemos

- Por qué el universo temprano estaba en un estado de entropía tan baja, cosa que el relato de la flecha por condiciones de contorno necesita pero no explica.
- Si el tiempo conserva un papel fundamental en una teoría cuántica de la gravitación satisfactoria o emerge de una estructura donde no es primitivo.
- Cómo se relaciona la experiencia psicológica del fluir temporal con las magnitudes físicas, hoy una pregunta sobre la cognición y no una que responda la física.

Qué podría hacernos cambiar de opinión

- Una comparación de relojes que no mostrara diferencia gravitatoria de ritmo a una separación de altura donde el desplazamiento fraccional predicho queda holgadamente dentro de la resolución demostrada, con los efectos sistemáticos controlados como en los experimentos con relojes ópticos, contradiría el relato del tiempo propio de la relatividad general.[3,5]
- Un experimento reproducible que identificara un sistema privilegiado en el que sucesos separados sean absolutamente simultáneos, compatible con los resultados nulos existentes, invalidaría la relatividad de la simultaneidad.[2]
- Una asimetría temporal demostrada en la dinámica fundamental capaz de dar cuenta de la flecha macroscópica observada, y no la asimetría temporal ya conocida en los procesos débiles que violan CP, eliminaría la necesidad de una explicación por condiciones de contorno.[6,7]

Si solo recuerdas una cosa

El tiempo en física es lo que acumulan los relojes a lo largo de su propio camino, y su dirección viene de cómo empezaron las cosas y no de las ecuaciones.

Artículos relacionados

- IDE-014 — Medición
- IDE-003 — Entropía
- IDE-023 — ¿Qué es la observación?

Referencias

- [1] Bureau International des Poids et Mesures. The International System of Units (SI), 9th edition, updated 2026. DOI 10.59161/AUEZ1291.
- [2] Albert Einstein. On the Electrodynamics of Moving Bodies. Annalen der Physik 17, 891-921 (1905). DOI 10.1002/andp.19053221004.
- [3] Charles W. Misner; Kip S. Thorne; John Archibald Wheeler. Gravitation. W. H. Freeman (1973), chapters 1 and 25.
- [4] R. V. Pound; G. A. Rebka Jr. Apparent Weight of Photons. Physical Review Letters 4, 337-341 (1960). DOI 10.1103/PhysRevLett.4.337.
- [5] C. W. Chou; D. B. Hume; T. Rosenband; D. J. Wineland. Optical Clocks and Relativity. Science 329, 1630-1633 (2010). DOI 10.1126/science.1192720.
- [6] Huw Price. Time's Arrow and Archimedes' Point: New Directions for the Physics of Time. Oxford University Press (1996).
- [7] Particle Data Group. CP Violation in the Quark Sector, in Review of Particle Physics, 2024 edition, revised August 2023 by T. Gershon and Y. Nir. <https://pdg.lbl.gov/2024/reviews/rpp2024-rev-cp-violation.pdf>

Medición: de la indicación al valor

IDE-014 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.6 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Práctica metrológica, del instrumento de laboratorio a los patrones nacionales

Límite: La medición como proceso operativo: mensurando, calibración, trazabilidad, incertidumbre y el modelo que convierte una lectura en un valor. La observación como acto epistémico, y qué cuenta como observar, corresponden a IDE-023

Trabajos relacionados: IDE-015, IDE-017, IDE-023

Perfil epistémico: Conocimiento establecido

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Un resultado de medida no es un número que entrega un instrumento. Es un valor atribuido a una magnitud definida, junto con un enunciado de la dispersión de valores que razonablemente podrían atribuírsele, obtenido mediante un modelo que conecta la indicación del instrumento con esa magnitud. [1,2] Cada parte de esa frase hace trabajo. Definir el mensurando va primero, porque una ambigüedad ahí no la repara un instrumento mejor. La calibración y la trazabilidad conectan una indicación con patrones de referencia mediante una cadena documentada, y cada eslabón añade incertidumbre en vez de quitarla. Y la incertidumbre no es la diferencia desconocida entre el resultado y la verdad. Este artículo sigue un resultado desde la indicación hasta la decisión, y deja la teoría general de los modelos a IDE-015.

En 60 segundos

La metrología empieza definiendo el mensurando, la magnitud que se pretende medir, y la definición puede ser el factor limitante: pedir la longitud de una barra es incompleto mientras no se enuncien la temperatura, las condiciones de apoyo y qué cuenta como sus extremos.[1] Solo entonces importa el instrumento. La calibración relaciona la indicación del instrumento con valores realizados por un patrón, y la trazabilidad es la cadena documentada de esas comparaciones hasta una referencia, normalmente una realización del SI.[1,3] La trazabilidad no vuelve exacto un resultado: cada eslabón aporta incertidumbre, y el valor de la cadena está en que las aportaciones se conocen y pueden sumarse. La incertidumbre es un parámetro que caracteriza la dispersión de valores que razonablemente podrían atribuirse al mensurando, evaluada estadísticamente a partir de observaciones repetidas o por otros medios, y no es la desviación desconocida respecto de un valor verdadero.[2] Entre la indicación y el valor hay un modelo de medida, y los resultados dependen de él. Cuando un resultado sirve para decidir si algo es conforme, la incertidumbre debe entrar explícitamente en la regla de decisión y no informarse para luego ignorarse.[4]

Por qué importa esta pregunta

La medición es el punto donde una teoría se encuentra con un número, y casi toda disputa sobre números es en realidad una disputa sobre una de cuatro cosas: qué se estaba midiendo, contra qué se

comparó, qué supuestos convirtieron la lectura en un valor y cuánta confianza cabe tener. Tratar un resultado como número pelado oculta las cuatro. Vuelve además invisibles dos errores frecuentes: leer dos valores como un desacuerdo cuando sus mensurandos nunca fueron la misma magnitud pretendida, y declarar conformidad o no conformidad sin contar la incertidumbre, lo que convierte una medida en un veredicto que no puede sostener.

1. Definir primero el mensurando

El mensurando es la magnitud que se pretende medir, y especificarlo exige más que nombrar un tipo de magnitud.[1] La especificación debe incluir cuantas condiciones afecten materialmente al valor: la temperatura a la que se define una longitud, el plano en que se toma un diámetro, la especie química exacta cuya concentración se busca y no una familia de compuestos emparentados.

Donde la definición es incompleta, la variabilidad residual no es error del instrumento. La metrología le da nombre, incertidumbre de definición, y fija un suelo por debajo del cual ningún instrumento puede llevarte.[1] Es una idea genuinamente útil también fuera de la metrología, porque explica por qué ciertos desacuerdos no se resuelven con aparatos más precisos.

La consecuencia no es que resultados sobre magnitudes distintas no puedan compararse: la metrología define la comparabilidad para resultados de magnitudes de la misma naturaleza trazables a una referencia común, y por eso los resultados de una distancia Tierra Luna y de una distancia París Londres son comparables cuando ambos trazan al metro.[1] Lo que sí exige es que, antes de leer dos resultados como un desacuerdo sobre el mismo objetivo, las especificaciones de sus mensurandos estén alineadas lo bastante para que sean resultados sobre la misma magnitud pretendida. La temperatura superficial del mar medida a la profundidad de la toma de agua y medida en la piel son magnitudes distintas que comparten nombre, y conciliarlas es cuestión de definición antes que de corrección. El desacuerdo aparente entre métodos de medida puede surgir de mensurandos o definiciones operacionales diferentes.

En qué punto estamos

El mensurando debe especificarse con todas las condiciones que afectan a su valor, y la ambigüedad que quede fija un suelo que ningún instrumento supera. La trazabilidad a una referencia común es lo que hace comparables resultados de la misma naturaleza; la definición del mensurando resuelve si dos resultados versan sobre la misma magnitud pretendida. Fijada la magnitud, la pregunta siguiente es contra qué se compara la lectura del instrumento.

2. La trazabilidad es una cadena, no un certificado

La calibración establece una relación entre las indicaciones de un instrumento y los valores que proporcionan patrones de medida, en condiciones enunciadas y con las incertidumbres de ambos. [1] No ajusta el instrumento, aunque el ajuste suela seguirla: documenta una relación.

La trazabilidad metrológica es la propiedad de un resultado por la que puede relacionarse con una referencia mediante una cadena ininterrumpida y documentada de calibraciones, cada una de las cuales contribuye a la incertidumbre de medida.[1] La referencia suele ser una realización de una unidad del SI, y desde la revisión de 2019 todas las unidades del SI se definen fijando los valores numéricos de un conjunto de constantes, de modo que realizarlas es cuestión de física y no de custodia de un artefacto.[3]

La expresión cada una contribuye a la incertidumbre es la parte que sostiene el peso. Un resultado trazable no es un resultado certificado como correcto: es uno cuya distancia a la referencia se ha contabilizado. Un certificado de calibración que no informa de incertidumbre no es prueba de exactitud sino un eslabón que falta.

En qué punto estamos

La calibración documenta la relación entre indicación y patrón, y la trazabilidad es la cadena documentada e ininterrumpida de esas relaciones hasta una referencia, con cada eslabón sumando incertidumbre. El sentido de la cadena es que las aportaciones se conocen, no que desaparezcan. Eso convierte el significado de la propia incertidumbre en lo siguiente que hay que precisar.

3. La incertidumbre no es el error

La GUM define la incertidumbre de medida como un parámetro que caracteriza la dispersión de los valores que razonablemente podrían atribuirse al mensurando.[2] Es un enunciado sobre un conjunto de valores compatibles con lo que se hizo, no una estimación de cuánto dista el resultado de un valor verdadero, porque el valor verdadero es por construcción desconocido.

La evaluación se divide por método y no por naturaleza. La evaluación de tipo A usa el análisis estadístico de observaciones repetidas; la de tipo B usa todo lo demás, incluidos certificados de calibración, especificaciones del fabricante, datos de manual y el conocimiento del analista sobre el proceso.[2] Ambas dan incertidumbres típicas que se combinan del mismo modo, que es lo que permite sumar coherentemente componentes de orígenes muy distintos.

De ahí dos hábitos. Informar un resultado sin incertidumbre es informar un resultado incompleto, porque no se ha dicho nada sobre qué valores son compatibles con él. Y dar más cifras de las que sostiene la incertidumbre afirma una precisión que la medida no tiene, lo que es una afirmación sobre el mundo y no una cuestión de formato.

En qué punto estamos

La incertidumbre caracteriza los valores que razonablemente pueden atribuirse al mensurando, evaluados a partir de observaciones repetidas o de otra información, y no es la distancia desconocida entre el resultado y la verdad. Un resultado sin ella está incompleto. Tanto la incertidumbre como el valor dependen de algo que hasta aquí ha quedado implícito: el modelo que conecta indicación y magnitud.

4. Todo resultado está mediado por un modelo

Entre la indicación del instrumento y el valor informado hay un modelo de medida: el conjunto de relaciones que expresan el mensurando en función de las magnitudes de entrada realmente observadas, junto con las correcciones por efectos conocidos.[2] Un termopar da una tensión; la temperatura sale de esa tensión más una corrección de la unión de referencia, más una relación de calibración, más supuestos sobre con qué está en contacto el sensor.

El relato basado en modelos vuelve el modelo central y no accesorio: un resultado de medida es una inferencia a partir de indicaciones, y su exactitud se evalúa por la robustez entre las consecuencias de modelos que representan procesos de medida distintos y no por comparación con un valor verdadero inaccesible.[7] Es una descripción de lo que hacen realmente los laboratorios de patrones cuando comparan realizaciones independientes de la misma unidad.

La consecuencia práctica es que los supuestos del modelo pertenecen al presupuesto de incertidumbre. Si se aplicó una corrección, su propia incertidumbre se propaga; si se despreció un efecto conocido por juzgarlo pequeño, ese juicio es también una entrada. Un resultado es por tanto condicional a un modelo enunciado, y cambiar el modelo puede cambiar el valor y su incertidumbre sin ninguna observación nueva.

En qué punto estamos

Un valor informado es una inferencia desde indicaciones a través de un modelo de medida, y los supuestos y correcciones del modelo pertenecen al presupuesto de incertidumbre y no a su exterior. La exactitud se evalúa por coherencia entre realizaciones independientes y no contra una verdad inaccesible. Queda qué ocurre cuando ese resultado debe sostener una decisión.

5. Del valor a la decisión

Casi toda medida existe para sostener una decisión: si esta pieza es conforme, si esta concentración está por debajo de un límite, si se ha alcanzado un objetivo. Un valor medido cerca de un límite no puede responder por sí solo, porque los valores compatibles con la medida quedan a ambos lados del límite.[4]

La evaluación de la conformidad exige por tanto una regla de decisión explícita. Las bandas de guarda estrechan el intervalo de aceptación respecto de la especificación para controlar el riesgo de aceptar un elemento no conforme, y la anchura de la banda se sigue de la incertidumbre y de qué error se considera peor.[4] Tanto la regla como el riesgo que acepta son elecciones, y deben enunciarse.

Aquí es donde toda la cadena se gana su sitio. La definición del mensurando fija sobre qué se decide; la trazabilidad hace significativa la comparación más allá de un laboratorio; la incertidumbre da a la regla de decisión con qué trabajar; y el modelo registra qué se supuso. Un número citado sin todo eso no es un resultado de medida sino una indicación con pretensiones.

En qué punto estamos

Un valor cercano a un límite no decide por sí solo la conformidad, así que hacen falta reglas de decisión y bandas de guarda construidas con la incertidumbre, y el riesgo que aceptan es una elección enunciada. Mensurando, trazabilidad, incertidumbre y modelo hacen trabajo en esa cadena. IDE-015 recoge los propios modelos y qué hace a uno adecuado para la inferencia que se le pide.

Lo que sabemos y cómo lo sabemos

- Definir el mensurando precede a la instrumentación, y la ambigüedad residual de la definición fija un suelo de incertidumbre que ningún instrumento rebaja.[1]
- La trazabilidad metrológica es una cadena documentada e ininterrumpida de calibraciones hasta una referencia en la que cada eslabón contribuye a la incertidumbre;[1] las referencias del SI se realizan hoy desde constantes fijadas y no desde artefactos.[3]
- La incertidumbre de medida caracteriza la dispersión de valores razonablemente atribuibles al mensurando y se evalúa por análisis estadístico de observaciones repetidas o por otros medios, que se combinan coherentemente.[2]
- Un valor informado es una inferencia desde indicaciones a través de un modelo de medida, así que los supuestos del modelo y las correcciones aplicadas pertenecen al presupuesto de incertidumbre.[2]

Lo que todavía no sabemos

- Cómo expresar mejor la incertidumbre de medida para magnitudes definidas por procedimientos complejos y no por relaciones físicas, cuestión viva en las ciencias biológicas y sociales.[6]
- Si el marco metrológico se transfiere, y cómo, a constructos medidos con cuestionarios y escalas de valoración, donde el mensurando lo define el propio instrumento.[6]
- Cómo tratar la incertidumbre que aporta un modelo computacional incrustado en una cadena de medida, donde la propagación no es analítica y el modelo puede ser opaco.

Qué podría hacernos cambiar de opinión

- Un caso demostrado en el que dos laboratorios que realizan la misma unidad del SI a partir de las constantes definitorias, con presupuestos de incertidumbre declarados por completo, discreparan mucho más que sus incertidumbres declaradas y la discrepancia resistiera toda resolución indicaría un defecto en las definiciones o en el marco de incertidumbre.[2,3]
- Un desarrollo que mostrara que un resultado de medida puede ser completo sin modelo de medida, de modo que una indicación se aplique a un valor sin paso inferencial, invalidaría el relato mediado por modelos.[5,7]
- Evidencia de que las reglas de decisión con bandas de guarda construidas desde incertidumbres declaradas estiman mal de forma sistemática el riesgo de aceptación en un entorno de producción bien caracterizado obligaría a revisar el marco de evaluación de la conformidad.[4]

Si solo recuerdas una cosa

Un resultado de medida es un valor con su incertidumbre para una magnitud definida, obtenido mediante un modelo y una comparación trazable; un número pelado no es nada de eso.

Artículos relacionados

- IDE-015 — Modelo
- IDE-017 — Incertidumbre
- IDE-023 — ¿Qué es la observación?

Referencias

- [1] JCGM 200:2012. International Vocabulary of Metrology: Basic and General Concepts and Associated Terms (VIM), 3rd edition. BIPM (2012).
- [2] JCGM 100:2008. Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement (GUM). BIPM (2008).
- [3] Bureau International des Poids et Mesures. The International System of Units (SI), 9th edition, V4.01, June 2026. DOI 10.59161/AUEZ1291.
- [4] JCGM 106:2012. Evaluation of Measurement Data: The Role of Measurement Uncertainty in Conformity Assessment. BIPM (2012).
- [5] Eran Tal. Making Time: A Study in the Epistemology of Measurement. The British Journal for the Philosophy of Science 67(1), 297-335 (2016). DOI 10.1093/bjps/axu037.
- [6] Luca Mari; Mark Wilson; Andrew Maul. Measurement across the Sciences: Developing a Shared Concept System for Measurement. Springer (2021). DOI 10.1007/978-3-030-65558-7.
- [7] Eran Tal. Measurement in Science. The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), Edward N. Zalta (ed.), sections 7.1 and 8.3.

Modelo: una máquina de inferencia con dominio

IDE-015 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.8 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Modelos analíticos, estadísticos y computacionales en todas las ciencias

Límite: Los modelos como máquinas de inferencia y cómo se evalúan. La relación de sustitución en sí corresponde a IDE-012; el modelo de medida dentro de una cadena metrológica corresponde a IDE-014

Trabajos relacionados: IDE-016, IDE-012, IDE-014

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Un modelo es una estructura deliberadamente simplificada construida para sostener inferencias sobre otra cosa. Su valor nunca es absoluto: es condicional a una pregunta, un dominio y un conjunto de supuestos, y un modelo excelente para una inferencia puede no servir para otra construida sobre el mismo sistema.[5,6] El aforismo de Box de que todos los modelos son falsos pero algunos son útiles se cita sin parar y se lee sin cuidado; lo que sostuvo fue que el científico debe preguntarse si un modelo ilumina y no si es verdadero, y atender a los supuestos gravemente equivocados y no a la falsedad universal.[1] Este artículo expone qué debe declarar un modelo, por qué el acuerdo con los datos es evidencia más débil de lo que parece, cómo se eligen modelos cuando varios ajustan, y qué hacer cuando modelos buenos discrepan.

En 60 segundos

Todo modelo omite, y lo que omite lo elige la pregunta para la que se construyó, así que lo primero que un modelo debe declarar es su propósito y su dominio.[6] Ajustar los datos es la evidencia más débil que puede ofrecer. Un modelo numérico de un sistema natural abierto no puede verificarse en sentido lógico, porque el acuerdo con las observaciones es compatible con muchas estructuras de modelo y con errores que se compensan; Oreskes y colaboradores sostienen que tales modelos pueden confirmarse en cierto grado y nunca establecerse como verdaderos.[2] La parsimonia tiene una versión precisa y no estética: los criterios de información estiman la pérdida predictiva esperada y penalizan parámetros porque la flexibilidad innecesaria degrada la predicción sobre datos nuevos, no porque la simplicidad sea elegante.[3,4] La inferencia multimodelo evalúa conclusiones sobre un conjunto de modelos plausibles y puede identificar las que permanecen estables en ese conjunto.[4] Y un modelo que ajusta por razones equivocadas, con parámetros que absorben el efecto de un mecanismo omitido, fallará justo donde se extrapole.

Por qué importa esta pregunta

Los modelos son el lugar donde de hecho se hacen las afirmaciones científicas, y los fallos más frecuentes no son errores dentro del modelo sino errores sobre él: usarlo fuera de su dominio, tomar el acuerdo con los datos por prueba y olvidar qué mecanismos quedaron fuera. Cada uno de ellos es invisible si el propósito y los supuestos del modelo no se enunciaron de entrada. Lo que está en juego es práctico, porque las decisiones de política e ingeniería descansan en extrapolaciones, y una extrapolación es precisamente el uso para el que la evidencia del ajuste pasado es más débil.

1. Un modelo se construye para una pregunta

Un modelo conserva algunos rasgos de su objetivo y descarta otros. Eso no es un compromiso lamentable sino la fuente de su utilidad: una representación que lo conservara todo sería tan inmanejable como el propio sistema. Lo que vuelve defendible la selección es que se hizo respecto de una pregunta, de modo que los rasgos omitidos son aquellos de los que la pregunta no depende. [5,6]

La representación y la similitud entre modelo y objetivo dependen del propósito.[5] Dos personas pueden escribir las mismas ecuaciones y estar afirmando cosas distintas, porque han asignado papeles distintos a los mismos términos.

La forma práctica de esto es una declaración. Un modelo debe enunciar para qué sirve, qué representa, qué excluye deliberadamente y en qué rango de condiciones se espera que valga. Un modelo sin eso no es más general: es menos comprobable, porque no hay condición enunciada bajo la cual pudiera detectarse que se aplica mal.

En qué punto estamos

Declarar propósito, exclusiones y dominio es lo que hace detectable una mala aplicación. Esa declaración cambia además lo que significa realmente la frase conocida sobre que todos los modelos son falsos.

2. Falso pero útil, dicho con precisión

El artículo de Box de 1976 suele reducirse a cinco palabras, y el argumento que las rodea es más útil que el eslogan. Escribiendo sobre modelos estadísticos de fenómenos naturales, sostuvo que, siendo falsos todos esos modelos, el científico no puede obtener uno correcto por elaboración excesiva y debe buscar en cambio una descripción económica que capture los rasgos importantes.[1] La recomendación es buscar los supuestos gravemente equivocados y no constatar que todos lo están.

Conviene conservar ese alcance. El asunto de Box son los modelos empíricos idealizados y no toda estructura formal, y leído como afirmación irrestricta de que nada puede describirse nunca correctamente el aforismo se convierte en una tesis filosófica y no en la metodológica que él defendió. Dentro de su alcance la distinción hace trabajo real: la falsedad universal no autorizaría nada, porque sería igual de cierta para el mejor y para el peor modelo disponible, mientras que si una simplificación concreta es consecuente para una inferencia concreta es una pregunta con respuesta, y la respuesta se halla perturbando el supuesto y mirando si la conclusión se mueve.

Box subrayó también la iteración. Un modelo se critica contra los datos, se diagnostican sus fallos y se modifica, lo que convierte la construcción de modelos en un proceso y no en un producto.[1] Los residuos son la parte productiva de ese lazo, porque un patrón sistemático en lo que el modelo no explicó es información sobre lo que quedó fuera.

En qué punto estamos

El argumento de Box, sobre modelos estadísticos de fenómenos naturales, es que ninguna elaboración produce uno correcto, así que el trabajo consiste en hallar qué supuestos están equivocados de forma consecuente para la inferencia en curso, e iterar contra los residuos. La falsedad universal por sí sola no excusa nada. La pregunta siguiente es qué establece realmente el acuerdo con los datos una vez ajustado un modelo a ellos.

3. Ajustar no es confirmar

Oreskes, Shrader-Frechette y Belitz sostuvieron que los modelos numéricos de sistemas naturales abiertos no pueden verificarse ni validarse en el sentido que esas palabras tienen en lógica y en ingeniería del software. Los sistemas naturales nunca están cerrados, los resultados del modelo nunca son únicos y el acuerdo con la observación es compatible con más de una estructura de modelo.[2]

El mecanismo es fácil de enunciar. Los parámetros se ajustan para encajar; un parámetro puede absorber el efecto de un mecanismo ausente; y dos errores pueden cancelarse en el rango de los datos. El modelo ajustado reproduce entonces las observaciones mientras codifica un relato equivocado de por qué salieron así, cosa indetectable desde el ajuste y visible solo fuera del rango ajustado.

Lo que tales modelos sí pueden es confirmarse en cierto grado por el número y la variedad de observaciones que reproducen, y usarse heurísticamente para explorar consecuencias de supuestos.[2] Es un estatuto epistémico real y no una descalificación, pero más débil de lo que sugiere el vocabulario de la validación, y la diferencia importa sobre todo en la extrapolación, que es el uso que la evidencia menos sostiene.

En qué punto estamos

El acuerdo con los datos no verifica un modelo de un sistema abierto, porque los parámetros pueden absorber mecanismos ausentes y los errores pueden cancelarse en el rango ajustado. La confirmación admite grados y proviene de la variedad de observaciones reproducidas, no de la bondad del ajuste. Queda la cuestión práctica de cómo elegir cuando varios modelos ajustan aceptablemente.

4. Elegir entre modelos

Añadir parámetros siempre mejora el ajuste a los datos disponibles, así que el ajuste no puede ser el criterio. La idea de Akaike fue definir el objetivo como el rendimiento predictivo esperado sobre datos nuevos del mismo proceso, y mostrar que la verosimilitud maximizada es un estimador sesgado de él, con un sesgo relacionado con el número de parámetros estimados.[3] El criterio resultante penaliza la complejidad por una razón enunciada y no por principio.

Eso convierte la parsimonia de preferencia estética en argumento predictivo. Un modelo con flexibilidad innecesaria ajusta el ruido igual que la señal, y sus predicciones sobre datos nuevos empeoran por ello. Burnham y Anderson desarrollan el aparato práctico, incluida la comparación de modelos candidatos sobre un conjunto común de datos y la ponderación de modelos por su apoyo relativo.[4]

Dos cautelas acompañan a la herramienta. Los criterios comparan los candidatos ofrecidos y no dicen nada sobre un modelo mejor ausente del conjunto, así que un modelo bien apoyado puede ser el mejor de un campo pobre. Y la comparación solo es válida entre modelos ajustados exactamente a los mismos datos, requisito fácil de violar sin darse cuenta.[4]

En qué punto estamos

La elección de modelo no puede apoyarse en el ajuste, porque el ajuste siempre mejora con parámetros, así que los criterios estiman el rendimiento esperado sobre datos nuevos y penalizan la complejidad por esa razón. La parsimonia pasa a ser un argumento predictivo y no un gusto. Los criterios ordenan solo los candidatos ofrecidos, lo que plantea qué hacer cuando varios modelos plausibles sobreviven juntos.

5. Cuando modelos buenos discrepan

La inferencia multimodelo evalúa conclusiones sobre un conjunto de modelos plausibles y puede identificar las que permanecen estables en ese conjunto. Una simulación molecular, una ecuación de medio continuo y un sustituto estadístico pueden acertar cada uno en su dominio sin ser intercambiables, porque responden preguntas distintas a costes distintos.[5,6] La discrepancia entre ellos informa en vez de avergonzar.

Eso da también la manera honesta de comunicar un resultado de modelado. Enuncia la pregunta, el dominio, los supuestos que se variaron, las conclusiones que fueron estables y las que no. Un solo número de un solo modelo, presentado sin esa estructura, es la forma en que se usan mal casi todos los resultados de modelado.

En qué punto estamos

Un resultado que se invierte al cambiar una elección estructural es un hecho sobre el modelo. IDE-016 recoge una propiedad que a los modelos se les pide cuantificar y que resiste una medida única: la complejidad.

Lo que sabemos y cómo lo sabemos

- La representación y la similitud entre modelo y objetivo dependen del propósito.[5,6]
- Box sostuvo, sobre modelos estadísticos de fenómenos naturales, que la elaboración no puede producir uno correcto y que la tarea es hallar qué supuestos están equivocados de forma consecuente, procediendo iterativamente contra los residuos.[1]
- Los modelos numéricos de sistemas abiertos no pueden verificarse; el acuerdo con los datos es compatible con múltiples estructuras y con errores que se compensan, así que la confirmación admite grados.[2]
- Los criterios de información penalizan parámetros porque la flexibilidad innecesaria degrada la predicción sobre datos nuevos, y ordenan solo los modelos candidatos realmente ofrecidos. [3,4]

Lo que todavía no sabemos

- Cómo caracterizar la fiabilidad predictiva de modelos de dimensión muy alta cuyo número efectivo de parámetros no está bien definido, donde los argumentos clásicos de penalización no se aplican sin más.
- Cómo ponderar entre sí modelos estructuralmente distintos cuando no se ajustan a una verosimilitud común, que es la situación habitual en los conjuntos multimodelo climáticos y epidemiológicos.
- Si la robustez en un conjunto de modelos plausibles es evidencia sobre el sistema o evidencia sobre los supuestos compartidos de la comunidad que produjo el conjunto.

Qué podría hacernos cambiar de opinión

- Un procedimiento demostrado que estableciera un modelo numérico de un sistema natural abierto como verdadero y no como confirmado en cierto grado, superando las objeciones de no unicidad y de errores que se compensan, invalidaría el estatuto epistémico estándar de esos modelos.[2]
- Evidencia reproducible de que los modelos seleccionados por criterios de información predicen peor datos nuevos del mismo proceso que los seleccionados por ajuste dentro de muestra, en entornos variados y con conjuntos de datos correctamente emparejados, socavaría la justificación predictiva de la parsimonia.[3,4]
- Un caso trabajado en el que una conclusión robusta en toda estructura de modelo plausible resultara falsa sobre el sistema, con el fallo atribuido a un supuesto compartido y no al azar, debilitaría el análisis de robustez como evidencia.[4,5]

Si solo recuerdas una cosa

Un modelo es una máquina de inferencia con pregunta, dominio y exclusiones enunciados; ajustar los datos es la evidencia más débil que puede ofrecer.

Artículos relacionados

- IDE-016 — Complejidad
- IDE-012 — Representación
- IDE-014 — Medición

Referencias

- [1] George E. P. Box. Science and Statistics. Journal of the American Statistical Association 71(356), 791-799 (1976). DOI 10.1080/01621459.1976.10480949.
- [2] Naomi Oreskes; Kristin Shrader-Frechette; Kenneth Belitz. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263, 641-646 (1994). DOI 10.1126/science.263.5147.641.
- [3] Hirotugu Akaike. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19(6), 716-723 (1974). DOI 10.1109/TAC.1974.1100705.
- [4] Kenneth P. Burnham; David R. Anderson. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer (2002).
- [5] Michael Weisberg. Simulation and Similarity: Using Models to Understand the World. Oxford University Press (2013).
- [6] Ronald N. Giere. How Models Are Used to Represent Reality. Philosophy of Science 71(5), 742-752 (2004). DOI 10.1086/425063.

Complejidad: varias preguntas bajo una palabra

IDE-016 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.5 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Cadenas, algoritmos, redes y ecosistemas

Límite: La complejidad como familia de medidas y qué mide cada una. La organización colectiva novedosa corresponde a IDE-007; los límites al conocimiento y la predicción corresponden a IDE-017

Trabajos relacionados: IDE-017, IDE-007, IDE-006

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

No existe una única magnitud llamada complejidad. Hay decenas de medidas y responden preguntas distintas: cuán larga debe ser una descripción, cuánto cómputo cuesta una respuesta, cuánta historia codifica una estructura, con qué fuerza dependen las partes unas de otras.[1] Y se separan de maneras que importan. Una cadena aleatoria es máximamente incompresible y no tiene ninguna organización, así que la longitud de descripción mide aleatoriedad y no estructura, y cualquier medida que alcance su máximo en el ruido está midiendo lo que no es para casi cualquier propósito.[2,4] Este artículo separa las familias principales, muestra por qué la medida debe elegirse con la pregunta, y cierra con un resultado que invirtió una intuición: añadir conexiones a un sistema grande puede volverlo menos estable, no más.[5]

En 60 segundos

Pregunta qué hace complejo a algo y obtendrás varias respuestas incompatibles, todas defendibles. La incompresibilidad algorítmica caracteriza la aleatoriedad; la profundidad lógica distingue por separado los objetos aleatorios o simples y someros de los objetos con historias computacionales profundas.[2] La complejidad computacional no pregunta cuán larga es la descripción sino cuánto trabajo cuesta la respuesta al crecer el problema, y la clasificación de problemas por ese coste alberga la pregunta abierta más difícil del campo.[3] La profundidad lógica pregunta una tercera cosa: cuánto cómputo haría falta para generar el objeto desde su descripción más corta, alta para estructuras que exigieron un proceso largo y baja tanto para cristales como para ruido.[4] La complejidad efectiva separa la parte regular de una descripción de su parte aleatoria y mide solo la primera.[6] Cada una es una magnitud real; ninguna es la complejidad, y usar una donde se quiere otra es la fuente habitual de confusión.

Por qué importa esta pregunta

Se invoca la complejidad para explicar por qué algo es difícil de predecir, difícil de construir o resistente a una descripción simple, y esas son propiedades distintas que pueden darse por separado. Una proteína

y una baraja barajada no son complejas en el mismo sentido, y afirmar que un sistema es demasiado complejo para modelarlo está vacío mientras no se diga qué medida es grande y por qué eso bloquea la inferencia que se intenta. La palabra arrastra además una implicación falsa que ha hecho daño real en ecología y en diseño de sistemas: que más partes y más conexiones compran robustez.

1. Ningún escalar único

La lista de medidas de Lloyd se titula deliberadamente no exhaustiva, y su interés es estructural y no enciclopédico. Las medidas se agrupan en torno a tres preguntas: cuán difícil es describir algo, cuán difícil es crearlo y en qué grado está organizado.[1] Son preguntas distintas, y ningún argumento muestra que deban tener una respuesta común.

Esto importa porque las afirmaciones sobre complejidad suelen colar una comparación. Decir que una célula es más compleja que un cristal es defendible bajo casi cualquier medida; decir que un ecosistema es más complejo que otro suele exigir nombrar la medida antes de significar algo. Inferencia: una afirmación sobre complejidad está incompleta si no identifica la medida utilizada.

La exigencia práctica es por tanto modesta y rara vez se cumple. Enuncia qué magnitud se afirma grande, sobre qué descripción del sistema y por qué esa magnitud es la relevante para la conclusión que se extrae. Casi todos los usos de la palabra fallan ya en la primera cláusula.

En qué punto estamos

Las medidas de complejidad forman familias distintas organizadas en torno a preguntas diferentes, entre ellas la dificultad de descripción, la dificultad de creación y el grado de organización.[1] La primera familia, la longitud de descripción, tiene una versión formal cuyo comportamiento resulta instructivo y al principio sorprendente.

2. La longitud de descripción y su trampa

Kolmogorov definió la complejidad de una cadena como la longitud del programa más corto que la produce en una máquina universal, lo que precisa la idea de que un objeto regular tiene una especificación corta.[2] Una cadena de un millón de ceros se comprime a una instrucción breve; una cadena sin patrón aprovechable, no.

La definición tiene dos propiedades conocidas que condicionan todo lo construido sobre ella. Depende de la máquina universal elegida, aunque solo hasta una constante aditiva, así que comparar cadenas largas tiene sentido y comparar cortas no.[2] Y no es computable: ningún procedimiento general halla el programa más corto, de modo que la magnitud es una referencia teórica y no algo que se mida.[7]

La incompresibilidad algorítmica caracteriza la aleatoriedad; la profundidad lógica distingue por separado los objetos aleatorios o simples y someros de los objetos con historias computacionales profundas.[2,4,6]

En qué punto estamos

La incompresibilidad algorítmica caracteriza la aleatoriedad; la profundidad lógica distingue por separado los objetos aleatorios o simples y someros de los objetos con historias computacionales profundas. Es además incomputable y dependiente de la máquina salvo una constante. Las medidas que evitan alcanzar su máximo en el ruido tienen que preguntar por algo distinto de la longitud de una descripción.

3. Esfuerzo e historia, no solo longitud

La complejidad computacional pregunta cómo crecen los recursos necesarios para resolver un problema con su tamaño. El resultado de Cook de 1971 identificó una clase de problemas a la que puede reducirse todo problema de una familia amplia, lo que permitió hablar de los problemas más difíciles de esa familia y no de casos difíciles sueltos.[3] Si esa familia admite solución eficiente es la pregunta abierta central del campo.

La profundidad lógica de Bennett pregunta otra cosa: cuánto cómputo hace falta para producir el objeto desde su descripción más corta.[4] Un cristal es superficial porque un programa breve lo produce deprisa; el ruido aleatorio es superficial porque el programa más corto es la propia cadena, impresa de inmediato. Profundo es un objeto cuya descripción breve existe pero exige un cómputo largo para desplegarse, imagen formal razonable de algo que necesitó una historia larga.

La complejidad efectiva de Gell-Mann y Lloyd toma la vía complementaria. Separa una descripción en la parte que captura regularidades y la parte que es detalle incompresible, y mide solo la parte regular.[6] El ruido puntúa entonces bajo porque casi toda su descripción es la componente aleatoria, que es el comportamiento que exige la noción intuitiva.

En qué punto estamos

La complejidad computacional mide el coste de una respuesta, la profundidad lógica el cómputo necesario para desplegar una descripción corta, y la complejidad efectiva solo la parte regular de una descripción. Cada una esquivó de otro modo la trampa del ruido. Las tres, sin embargo, son relativas a cómo se describió el objeto de partida.

4. Complejidad ¿de qué, exactamente?

Las medidas de complejidad incluyen tipos basados en la descripción y tipos computacionales o generativos; la complejidad efectiva depende específicamente de qué regularidades se identifiquen. [1,6] Un gas se describe trivialmente con dos variables termodinámicas y resulta imposiblemente complicado como lista de posiciones moleculares, y ambas descripciones son del mismo gas.

La disciplina es la misma que en el resto de la serie. Nombra la descripción, nombra la medida y comprueba que la conclusión depende de la magnitud realmente calculada.

En qué punto estamos

Las medidas de complejidad incluyen tipos basados en la descripción y tipos computacionales o generativos; la complejidad efectiva depende específicamente de qué regularidades se identifiquen.[1,6] Eso afecta a una creencia extendida sobre qué compra la complejidad.

5. Más partes no es más robusto

Fue durante un tiempo supuesto de trabajo en ecología que las comunidades más diversas y más conectadas son más estables. May analizó comunidades modelo ensambladas al azar y encontró la tendencia contraria: al aumentar el número de especies, la conectancia y la intensidad de interacción, la probabilidad de que el equilibrio sea localmente estable cae con fuerza.[5]

El resultado trata de una clase concreta de modelos, matrices de interacción construidas al azar, y de la estabilidad local en un equilibrio, y May fue explícito al respecto. Su fuerza no está en que los ecosistemas reales sean inestables sino en que la estabilidad no puede inferirse solo de la

complejidad: lo que estabiliza las comunidades reales tiene que ser un rasgo de cómo están estructuradas y no de cuántas partes tienen.

La lección general se transfiere. Añadir componentes y acoplamientos a cualquier sistema agranda el espacio de maneras en que puede fallar, y la robustez viene de la arquitectura y no del recuento. Es la misma conclusión que llega desde el lado de la realimentación, donde la estabilidad depende de la disposición de ganancias y retardos y no del número de lazos.

En qué punto estamos

May mostró que en comunidades modelo ensambladas al azar un mayor tamaño, conectancia e intensidad de interacción reducen la probabilidad de estabilidad local, así que la complejidad no compra robustez y lo que estabiliza los sistemas reales es estructural. El resultado está acotado a su clase de modelos, y eso forma parte de leerlo bien. IDE-017 recoge la familia de límites que suele achacarse a la complejidad: la incertidumbre y sus fuentes.

Lo que sabemos y cómo lo sabemos

- Las medidas de complejidad forman familias distintas organizadas en torno a preguntas diferentes, entre ellas la dificultad de descripción, la dificultad de creación y el grado de organización.[1]
- La complejidad de Kolmogórov es la longitud del programa más corto que produce una cadena, depende de la máquina solo hasta una constante aditiva y es incomputable; la imprevisibilidad algorítmica caracteriza la aleatoriedad.[2,7]
- La profundidad lógica mide el cómputo necesario para generar un objeto desde su descripción más corta y es baja tanto para cristales como para ruido; la complejidad efectiva mide solo la parte regular de una descripción.[4,6]
- En comunidades modelo ensambladas al azar, aumentar el número de especies, la conectancia y la intensidad de interacción reduce la probabilidad de estabilidad local, así que la complejidad no confiere robustez por sí misma.[5]

Lo que todavía no sabemos

- Si la familia amplia de problemas identificada por la reducción de Cook admite solución eficiente, pregunta abierta central de la complejidad computacional.[3]
- Qué medida de complejidad, si alguna, sigue la noción intuitiva de estructura organizada a través de sistemas físicos, biológicos y sociales y no solo dentro de un dominio.
- Hasta dónde se extiende el resultado de estabilidad de May a estructuras de interacción no ensambladas al azar, que es donde viven las redes ecológicas y de ingeniería reales.

Qué podría hacernos cambiar de opinión

- Una demostración que resolviera si la familia de problemas identificada por Cook admite solución eficiente zanjaría la pregunta central del campo y reorganizaría la clasificación de la dificultad computacional.[3]
- Una medida computable que asignara valores bajos tanto a cadenas aleatorias como perfectamente regulares, valores altos a la estructura organizada, y coincidiera con la profundidad lógica y la complejidad efectiva donde estas son calculables, sustituiría la pluralidad actual de medidas.[1,4,6]

- Evidencia empírica de que las comunidades ecológicas con mayor conectancia son sistemáticamente más estables ante perturbaciones comparables, controlada la estructura de interacción, obligaría a revisar cómo se generaliza el resultado de May.[5]

Si solo recuerdas una cosa

La complejidad comprende familias distintas de medidas relativas a descripción, creación y organización.[1,6]

Artículos relacionados

- IDE-017 — Incertidumbre
- IDE-007 — Emergencia
- IDE-006 — Escala

Referencias

- [1] Seth Lloyd. Measures of Complexity: A Nonexhaustive List. IEEE Control Systems Magazine 21(4), 7-8 (2001). DOI 10.1109/37.969131.
- [2] A. N. Kolmogorov. Three Approaches to the Quantitative Definition of Information. Problems of Information Transmission 1(1), 1-7 (1965).
- [3] Stephen A. Cook. The Complexity of Theorem-Proving Procedures. Proceedings of the Third Annual ACM Symposium on Theory of Computing, 151-158 (1971). DOI 10.1145/800157.805047.
- [4] Charles H. Bennett. Logical Depth and Physical Complexity. In: Rolf Herken (ed.), The Universal Turing Machine: A Half-Century Survey, Oxford University Press, 227-257 (1988).
- [5] Robert M. May. Will a Large Complex System be Stable? Nature 238, 413-414 (1972). DOI 10.1038/238413a0.
- [6] Murray Gell-Mann; Seth Lloyd. Information Measures, Effective Complexity, and Total Information. Complexity 2(1), 44-52 (1996).
- [7] Ming Li; Paul Vitányi. An Introduction to Kolmogorov Complexity and Its Applications, 4th edition. Springer (2019), chapter 2.

Incertidumbre: nombrar las fuentes antes de sumarlas

IDE-017 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.5 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Del presupuesto de un instrumento a la proyección con relevancia política

Límite: La incertidumbre como conjunto de fuentes y cómo se separan, propagan y comunican. La probabilidad como cálculo corresponde a IDE-008; el presupuesto metrológico de incertidumbre de un resultado corresponde a IDE-014

Trabajos relacionados: IDE-018, IDE-008, IDE-014

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La incertidumbre no es una sola cosa con un número asociado. Una magnitud puede ser incierta porque el instrumento dispersa, porque un parámetro se estimó con datos limitados, porque el modelo omite un mecanismo o porque el curso futuro de los acontecimientos no está determinado. Esas fuentes se comportan de manera distinta: unas se reducen con más datos y otras no, y juntarlas en un solo intervalo oculta cuál es cuál. Este artículo separa las fuentes, sostiene que la división habitual entre aleatoria y epistémica depende de dónde trace el modelizador la frontera, muestra que las conclusiones sobre parámetros están condicionadas al modelo adoptado y a sus supuestos, y toma en serio los casos en que la cuantificación no está disponible.[3,4]

En 60 segundos

Empieza preguntando qué reduciría la incertidumbre. Si más medidas del mismo tipo la estrecharían, se comporta como un problema de estimación; si no, algo estructural es responsable. Ese criterio hace casi todo el trabajo que se supone hacen las etiquetas aleatoria y epistémica, y evita su trampa principal: que la variabilidad cuente como azar inherente o como conocimiento ausente depende del modelo elegido y no del mundo, así que dos analistas competentes pueden clasificar la misma magnitud de las dos maneras.[2] De ahí una segunda advertencia. Las conclusiones sobre sensibilidad y parámetros están condicionadas al modelo adoptado y a sus supuestos.[3] Por último, no todo es cuantificable. Knight distinguió el riesgo medible de la incertidumbre que no se deja reducir a una distribución, y fingir lo contrario produce números sin respaldo.[4,5]

Por qué importa esta pregunta

La incertidumbre es donde un resultado científico se encuentra con una decisión, y cómo se comunica determina qué decisión cabe. Un intervalo único invita a tratar toda la duda como de un solo tipo, lo que oculta las dos preguntas que importan para actuar: si más evidencia del mismo tipo ayudaría y qué tendría que estar mal para que la respuesta cambiara. Vuelve además invisible un fallo frecuente, en el

que un intervalo calculado con precisión alrededor de un parámetro se ofrece como si acotara el error de la conclusión, cuando la incertidumbre dominante está en la estructura que la produjo.

1. Las fuentes se comportan de manera distinta

Conviene enumerar las fuentes por separado porque responden a remedios distintos. La incertidumbre de medida procede de la calibración, la resolución, los efectos ambientales y el modelo de medida, y es la que tiene el tratamiento formal más desarrollado.[1] La incertidumbre de parámetros procede de estimar valores con datos finitos. La estructural procede de la elección de la forma del modelo y de los mecanismos omitidos. La de escenario procede de condiciones que no están determinadas en absoluto, como las decisiones futuras.

El diagnóstico útil es qué reduciría cada una. Más observaciones del mismo tipo reducen la incertidumbre de parámetros y, hasta el suelo definitorio, la de medida. No reducen la estructural, que exige otro modelo, y no hacen nada con la de escenario, que no es un estado de ignorancia sino una apertura genuina.

Informarlas juntas como un solo intervalo no es por tanto conservador: es poco informativo. Un resultado dado como un valor más menos un número invita al lector a suponer que ese número acota la duda, cuando puede acotar solo su componente más tratable.

En qué punto estamos

Las incertidumbres de medida, de parámetros, estructural y de escenario responden a remedios distintos, y la pregunta diagnóstica es qué reduciría cada una. Colapsarlas en un intervalo oculta cuál domina. El intento más común de ordenarlas es una división en dos, y esa división resulta menos absoluta de lo que parece.

2. Aleatoria y epistémica, y por qué la línea se mueve

La distinción estándar separa la incertidumbre aleatoria, tratada como variabilidad inherente, de la epistémica, tratada como conocimiento incompleto que en principio podría reducirse. Es una división de trabajo útil, y la práctica en ingeniería depende de ella porque ambas se propagan de forma distinta en un análisis de fiabilidad.

Der Kiureghian y Ditlevsen sostienen que la distinción es rasgo del modelo y no de la naturaleza. Una variable tratada como aleatoria en un modelo puede representarse explícitamente, y volverse por tanto epistémica, en otro más detallado; la clasificación depende entonces de la frontera que trazó el analista.[2] Lo que importa para una decisión, argumentan, no es la etiqueta sino si la incertidumbre puede reducirse con el tiempo y los recursos disponibles.

Ese replanteamiento es más útil que las etiquetas. Preguntar si un componente es reducible en la práctica, y a qué coste, produce una respuesta que guía la acción, mientras que preguntar si es aleatorio en principio produce una disputa sobre convenciones de modelado.

En qué punto estamos

La división entre aleatoria y epistémica es propiedad del modelo y no del mundo, ya que una variable tratada como aleatoria en una descripción puede representarse explícitamente en otra más fina. La pregunta que sobrevive es si un componente es reducible en la práctica y a qué coste. Una confusión emparentada y más peligrosa atañe a cuál componente domina de entrada.

3. La precisión de los parámetros está condicionada al modelo

Las conclusiones sobre sensibilidad y parámetros están condicionadas al modelo adoptado y a sus supuestos.[3]

La incertidumbre estructural es más difícil de cuantificar y a menudo mayor. Inferencia: bajo una especificación errónea, un parámetro ajustado puede compensar un mecanismo omitido, por lo que un intervalo estrecho no implica necesariamente un modelo adecuado.

El análisis global de sensibilidad es la herramienta estándar para averiguar de qué depende la respuesta. En vez de variar una entrada cada vez en torno a un punto nominal, explora conjuntamente el espacio de entradas y reparte la varianza de la salida entre las entradas y sus interacciones, que es lo que revela que un resultado lo impulsa un supuesto que nadie vigilaba.[3]

En qué punto estamos

El análisis global de sensibilidad es lo que identifica de qué supuestos descansa realmente la respuesta. Más allá de todo esto está el caso en que no hay distribución alguna.

4. Cuando la cuantificación no está disponible

Knight distinguió el riesgo medible, donde la distribución de resultados se conoce o puede estimarse, de la incertidumbre propiamente dicha, donde no.[4] La distinción se hizo sobre decisiones económicas pero su estructura es general: unas situaciones ofrecen una clase de referencia bien definida y otras no, y una probabilidad calculada sin ella es una afirmación y no una medida.

No es un argumento contra cuantificar. Es un argumento a favor de decir en qué caso estás. Una probabilidad informada que descansa en una frecuencia defendible o en un modelo explícitamente enunciado se gana su estatuto; una producida porque hacía falta un número tiene el mismo aspecto y ninguno de sus respaldos. El marco bayesiano vuelve explícitos los supuestos en vez de eliminarlos, lo que es una ganancia real, porque la distribución previa y el modelo quedan a la vista y pueden discutirse.[6]

Donde la cuantificación no está genuinamente disponible, las opciones honestas son describir el rango de posibilidades, identificar qué las distinguiría y buscar decisiones que rindan aceptablemente en todo el rango en vez de óptimamente bajo una distribución supuesta.

En qué punto estamos

La distinción de Knight entre riesgo medible e incertidumbre no cuantificable trata de si existe una clase de referencia, y una probabilidad producida sin ella es una afirmación y no una medida. Explicitar los supuestos es una ganancia aunque no los elimine. Queda cómo decir todo esto a quien tiene que actuar.

5. Decirlo con honestidad

Funtowicz y Ravetz propusieron que un resultado cuantitativo llevado a la política vaya acompañado de algo más que una barra de error: de una valoración de la calidad de la información subyacente, incluidos cómo se produjeron los números y cuánto del resultado descansa en juicio experto.[5] La evaluación de la incertidumbre incluye la calidad y el pedigrí de la información subyacente, no solo la anchura de un intervalo numérico. Inferencia: intervalos de igual anchura pueden descansar, por tanto, en evidencia de calidad diferente.

Sea cual sea la notación, tres cosas deben sobrevivir al resumen. Qué fuentes de incertidumbre se incluyeron y cuáles no; si el componente dominante es reducible con más del mismo trabajo; y qué

tendría que ser cierto para que la conclusión cambiara. Un resultado informado sin eso no es más prudente por llevar un intervalo adosado.

La tentación va en el otro sentido. Quien decide pide un número, y un intervalo que omite la incertidumbre estructural es más estrecho y parece más útil que uno que la incluye. Resistirlo es toda la disciplina, y resulta más fácil cuando las fuentes se separaron de entrada en vez de agregarse y defenderse después.

En qué punto estamos

La evaluación de la incertidumbre incluye la calidad y el pedigrí de la información subyacente, no solo la anchura de un intervalo numérico. Los intervalos más estrechos resultan más atractivos y suelen ser menos completos. IDE-018 recoge una propiedad estructural que determina cómo se propaga la incertidumbre por un sistema: cómo están conectadas sus partes.

Lo que sabemos y cómo lo sabemos

- Las incertidumbres de medida, de parámetros, estructural y de escenario responden a remedios distintos, y solo algunas se reducen con más observaciones del mismo tipo. Ese esquema de cuatro vías es la distinción organizadora propia de este artículo y no un resultado tomado de una fuente.
- La clasificación en aleatoria y epistémica depende de la frontera de modelo que trazó el analista y no de la naturaleza, así que la pregunta operativamente útil es si un componente es reducible en la práctica.[2]
- Las conclusiones sobre sensibilidad y parámetros están condicionadas al modelo adoptado y a sus supuestos.[3]
- El análisis global de sensibilidad explora conjuntamente el espacio de entradas y reparte la varianza de la salida entre entradas e interacciones, que es como se identifican los supuestos dominantes.[3]

Lo que todavía no sabemos

- Cómo caracterizar y propagar la incertidumbre estructural de forma fundamentada cuando el conjunto de estructuras de modelo plausibles no puede enumerarse.
- Si la incertidumbre profunda en el sentido de Knight puede representarse formalmente, y cómo, sin importar la precisión injustificada que la distinción se trazó para evitar.[4]
- Qué formatos de comunicación de la incertidumbre mejoran de hecho las decisiones, pregunta empírica sobre usuarios y no sobre matemáticas.

Qué podría hacernos cambiar de opinión

- Una demostración de que la incertidumbre estructural en una clase de modelos puede acotarse rigurosamente solo con datos, sin enumerar estructuras alternativas, eliminaría la razón principal por la que los intervalos de parámetros se consideran incompletos.[3]
- Evidencia de que analistas que clasifican de forma independiente las mismas incertidumbres como aleatorias o epistémicas coinciden en alto grado con independencia de sus elecciones de modelado contradiría la tesis de que la distinción es relativa al modelo.[2]
- Un caso trabajado en el que una probabilidad asignada sin clase de referencia defendible superara, en muchas repeticiones, a un enunciado explícito del rango de posibilidades, debilitaría el argumento contra cuantificar la incertidumbre profunda.[4,6]

Si solo recuerdas una cosa

La incertidumbre tiene fuentes que se comportan de manera distinta, y un intervalo que no dice cuáles incluyó oculta más de lo que informa.

Artículos relacionados

- IDE-018 — Conexión
- IDE-008 — Probabilidad
- IDE-014 — Medición

Referencias

- [1] JCGM 100:2008. Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement (GUM). BIPM (2008).
- [2] Armen Der Kiureghian; Ove Ditlevsen. Aleatory or epistemic? Does it matter? Structural Safety 31(2), 105-112 (2009). DOI 10.1016/j.strusafe.2008.06.020.
- [3] Andrea Saltelli; Marco Ratto; Terry Andres; Francesca Campolongo; Jessica Cariboni; Debora Gatelli; Michaela Saisana; Stefano Tarantola. Global Sensitivity Analysis: The Primer. Wiley (2008).
- [4] Frank H. Knight. Risk, Uncertainty and Profit. Houghton Mifflin (1921).
- [5] Silvio O. Funtowicz; Jerome R. Ravetz. Uncertainty and Quality in Science for Policy. Theory and Decision Library A, volume 15. Kluwer Academic Publishers (1990). ISBN 978-0-7923-0799-0.
- [6] Edwin T. Jaynes; edited by G. Larry Bretthorst. Probability Theory: The Logic of Science. Cambridge University Press (2003). ISBN 978-0-521-59271-0.

Conexión: qué afirma una arista

IDE-018 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguro

Versión: v0.6 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Redes sociales, biológicas, tecnológicas e infraestructurales

Límite: La estructura de conexión y qué determina. La dinámica de un lazo cerrado corresponde a IDE-010; qué transporta físicamente una influencia corresponde a IDE-009 e IDE-024

Trabajos relacionados: IDE-019, IDE-010, IDE-016

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Dibujar una red reduce un sistema a nodos y aristas, y cada arista es una afirmación: que estas dos cosas se relacionan de una manera especificada que importa para la pregunta que se hace. La abstracción es potente porque la arquitectura, y no el mero número de enlaces, determina cómo se comporta un sistema. Un poco de recableado hunde las longitudes de camino sin destruir el agrupamiento local; una regla de crecimiento que favorece a los ya conectados produce unos pocos concentradores muy grandes; y las redes de ese tipo toleran bien el fallo aleatorio y son frágiles ante la eliminación dirigida.[1,2,3] Este artículo expone esos resultados y expone también dónde el campo ha tenido que volverse más cuidadoso, porque las distribuciones de grado de cola pesada se informan más a menudo de lo que sostienen los datos.[4]

En 60 segundos

Una red vale lo que vale su definición de arista, porque el mismo conjunto de entidades da redes distintas según qué cuente como enlace.[6] Fijada la arista, la estructura hace trabajo real. Watts y Strogatz mostraron que recablear una fracción pequeña de los enlaces de una red regular hunde la longitud típica de camino dejando casi intacto el agrupamiento local, y por eso redes muy agrupadas pueden tener rutas cortas entre dos nodos cualesquiera.[1] Barabási y Albert mostraron que el crecimiento más una unión proporcional al grado existente produce una distribución de grado de cola pesada, lo que da un mecanismo para los concentradores, aunque no el único.[2] En redes de ese tipo, con grados esencialmente no correlacionados, los concentradores explican una asimetría llamativa: eliminar nodos al azar apenas perturba la conectividad, mientras que eliminar los de mayor grado fragmenta la red deprisa.[3] La cautela es que identificar una ley de potencias a partir de datos es más difícil de lo que parece, y muchas afirmaciones publicadas no sobreviven a una comprobación estadística adecuada.[4]

Por qué importa esta pregunta

El lenguaje de redes se ha extendido a todo campo con datos relacionales, y con él dos hábitos que conviene resistir. El primero es dibujar aristas sin decir qué significan, lo que produce una imagen que parece explicativa y no restringe nada. El segundo es leer un resultado estructural como ley universal: que deben existir concentradores, que las redes son robustas, que seis grados separan a todo el mundo.

Cada uno de ellos es un resultado condicional sobre estructuras particulares, y conocer las condiciones es lo que separa usar la herramienta de decorar con ella.

1. Una arista es una afirmación

Una representación de red afirma que el sistema puede describirse útilmente con un conjunto de entidades y un conjunto de relaciones por pares de un tipo enunciado. Ambas mitades son elecciones. Una representación de red depende de qué se considere vértice y arista; los pesos, las direcciones y la agregación temporal pueden alterar la red representada.[6]

Como la definición de arista es una elección, puede ser la equivocada para la pregunta. Una red de contactos construida con colocación a resolución de una hora es un objeto distinto de otra construida a resolución de un minuto, y una afirmación sobre propagación de una enfermedad puede valer en una y fallar en la otra.

La exigencia práctica es que una figura de red vaya acompañada de qué es un nodo, qué es una arista, cómo se determinó y qué se excluyó. Sin eso, el diagrama no puede estar equivocado, lo que significa que tampoco está haciendo trabajo científico.

En qué punto estamos

Un diagrama que no dice qué significa una arista no puede comprobarse. Fijada la arista, la primera sorpresa estructural es cuán poco recableado hace falta para cambiar las distancias de una red.

2. Caminos cortos sin cableado denso

Una red regular tiene alto agrupamiento local, porque los vecinos comparten vecinos, y caminos típicos largos, porque atravesarla exige muchos pasos cortos. Un grafo aleatorio tiene el perfil opuesto. Watts y Strogatz interpolaron entre ambos recableando cada arista con probabilidad pequeña y encontraron que las dos propiedades no se intercambian de forma suave.[1]

La longitud de camino cae con fuerza a una probabilidad de recableado donde el agrupamiento sigue casi en su valor de red regular. Un número muy pequeño de atajos de largo alcance basta para hundir la distancia media, de modo que hay un régimen amplio en el que una red está a la vez muy agrupada y tiene caminos cortos.[1] Ahí es donde se sitúan muchas redes reales.

El resultado explica algo que parecía paradójico. Las redes sociales son localmente densas, porque los amigos de uno se conocen entre sí, y sin embargo cadenas cortas conectan a desconocidos distantes. Los dos hechos no están en tensión, porque los atajos responsables de los caminos cortos son lo bastante raros como para no perturbar la estructura local.

En qué punto estamos

Recablear una fracción pequeña de las aristas de una red regular hunde la longitud típica de camino dejando casi intacto el agrupamiento, así que ser localmente denso y globalmente cercano no es una contradicción. Unos pocos atajos de largo alcance hacen el trabajo. La longitud de camino es una propiedad estructural; la distribución de cuántos enlaces tiene cada nodo es otra, y se comporta muy distinto.

3. Concentradores, y con qué frecuencia están de verdad

Barabási y Albert propusieron un relato generativo de las distribuciones de grado de cola pesada: deja que la red crezca añadiendo nodos y que cada nodo nuevo se una preferentemente a los que

ya tienen muchos enlaces. La dinámica de que el rico se enriquece produce un pequeño número de concentradores de grado muy alto y una cola larga.[2]

El modelo importa como mecanismo, y no debe leerse como la afirmación de que todas las redes son así ni de que la unión preferencial sea la única vía a una cola pesada. Otros procesos, incluidos el crecimiento multiplicativo simple y ciertas dinámicas de optimización, producen distribuciones parecidas, así que observar una cola pesada no identifica el mecanismo que la produjo.[4,6]

La cautela más fuerte es estadística. Clauset, Shalizi y Newman mostraron que ajustar una recta a un histograma log-log es una manera poco fiable de identificar una ley de potencias, y que al aplicar estimación y contrastes de bondad de ajuste adecuados muchos ejemplos publicados no quedan bien descritos por una ley de potencias, con alternativas log-normales o exponenciales estiradas ajustando igual o mejor.[4] La existencia de concentradores suele ser real; la forma funcional concreta con frecuencia no está establecida.

En qué punto estamos

La unión preferencial da un mecanismo para las distribuciones de grado de cola pesada, pero no el único, así que una cola pesada no identifica su causa. La estimación y el contraste adecuados muestran además que muchas leyes de potencias informadas no las sostienen los datos. Sea cual sea su origen, la presencia de concentradores tiene una consecuencia sobre el fallo que conviene enunciar con cuidado.

4. La arquitectura decide la robustez

Albert, Jeong y Barabási compararon dos tipos de eliminación de nodos en redes modelo con distribuciones de grado de cola pesada y en dos redes empíricas grandes. Eliminar nodos al azar deja la conectividad notablemente intacta, porque un nodo elegido al azar es casi con seguridad uno de los muchos de grado bajo. Eliminar nodos por orden de grado fragmenta la red tras retirar una fracción pequeña.[3] La distribución de grado no es toda la condición: el argumento supone grados esencialmente no correlacionados, y redes cuyos nodos de grado alto se unen preferentemente entre sí, o que son fuertemente modulares, pueden comportarse de otro modo a los tamaños realmente estudiados.

El mismo rasgo estructural produce por tanto la robustez y la fragilidad, y por eso ninguna de las dos palabras describe la red por sí sola. Una red es robusta o frágil respecto de una clase especificada de perturbación, y un sistema diseñado para una puede quedar muy expuesto a la otra.

Esto conecta directamente con un resultado que llega por otra vía. Añadir partes y acoplamientos no compra estabilidad por sí mismo, y lo que decide si un sistema grande se sostiene es cómo están dispuestos los acoplamientos. La literatura de redes y la de sistemas dinámicos llegan a esa conclusión de forma independiente.

En qué punto estamos

En las redes estudiadas, de cola pesada y con grados casi no correlacionados, el fallo aleatorio se tolera bien y la eliminación dirigida no, y la misma estructura produce ambas cosas, así que la robustez debe enunciarse respecto de una clase de perturbación y de una clase de red. Contar partes no predice nada; la arquitectura sí. Falta decir qué viaja por una arista y si todas transportan lo mismo.

5. Qué transportan las conexiones

Granovetter sostuvo que los lazos débiles, las relaciones infrecuentes y de poca intimidad, son desproporcionadamente valiosos para el flujo de información novedosa, porque los lazos fuertes tienden a conectar a personas cuyos contactos ya se solapan. Un lazo débil tiene más probabilidad de ser un puente hacia otra parte de la red y, por tanto, de transportar información que los fuertes no pueden.[5]

Cuando la dinámica de propagación se acopla a la heterogeneidad de los enlaces, el peso, la dirección o el tipo de arista pueden afectar a la propagación predicha; los lazos débiles y los puentes pueden desempeñar funciones estructurales distintas en la difusión de información.[5,6] Lo mismo vale en otros sitios: las aristas de una red metabólica representan reacciones con velocidades y direcciones, y una imagen que las ignore describe otro sistema.

Esto devuelve el artículo a donde empezó. Los resultados estructurales son fuertes, pero todos ellos son condicionales a qué significan las aristas y a qué se supone que fluye por ellas. Enunciarlo no es una advertencia añadida a un análisis de redes: es la parte que hace que el análisis trate de algo.

En qué punto estamos

Cuando la dinámica de propagación se acopla a la heterogeneidad de los enlaces, el peso, la dirección o el tipo de arista pueden afectar a la propagación predicha; los lazos débiles y los puentes pueden desempeñar funciones estructurales distintas en la difusión de información. Todo resultado estructural es condicional a qué significan las aristas. IDE-019 pasa a algo que viaja entre lugares separados por derecho propio, donde hay que distinguir portador, propagación y detección: la luz.

Lo que sabemos y cómo lo sabemos

- Una representación de red depende de qué se considere vértice y arista; los pesos, las direcciones y la agregación temporal pueden alterar la red representada.[6]
- Recablear una fracción pequeña de las aristas de una red regular hunde la longitud típica de camino dejando casi intacto el agrupamiento, así que alto agrupamiento y caminos cortos coexisten en un régimen amplio.[1]
- El crecimiento con unión proporcional al grado existente genera distribuciones de grado de cola pesada, aunque otros procesos generan distribuciones parecidas y muchas leyes de potencias informadas no sobreviven a un contraste estadístico adecuado.[2,4]
- En redes de cola pesada con grados esencialmente no correlacionados, la eliminación aleatoria de nodos se tolera bien y la ordenada por grado fragmenta la red deprisa, así que una misma estructura da robustez y fragilidad; la distribución de grado por sí sola no fija el comportamiento.[3]

Lo que todavía no sabemos

- Qué mecanismo generativo produce las distribuciones de grado de cola pesada observadas en una red real concreta, dado que varios mecanismos dan distribuciones parecidas.[2,4]
- Cuánta de la universalidad estructural informada sobrevive una vez estandarizadas entre estudios las definiciones de arista, el muestreo y los umbrales.[4,6]
- Cómo caracterizar la robustez en redes cuyas aristas transportan cosas cualitativamente distintas, donde una sola medida de conectividad no captura qué puede fallar.

Qué podría hacernos cambiar de opinión

- Un reanálisis sistemático que mostrara que distribuciones empíricas de grado rechazadas como leyes de potencias por contrastes basados en verosimilitud sí quedan bien descritas por ellas bajo un tratamiento correcto del muestreo invertiría la cautela estadística actual.[4]
- Una clase demostrada de redes de cola pesada y grados esencialmente no correlacionados, del tipo y tamaño que cubre el resultado, que se fragmentara bajo eliminación aleatoria con la misma facilidad que bajo eliminación por grado contradiría el resultado de tolerancia a errores y ataques.[3]
- Evidencia de que alto agrupamiento y longitud característica corta de camino no pueden coexistir en un régimen amplio de parámetros, en contra del resultado de interpolación, eliminaría la explicación estructural del comportamiento de mundo pequeño.[1]

Si solo recuerdas una cosa

Una red es una afirmación sobre qué cuenta como enlace, y su arquitectura, no su número de enlaces, decide cómo se comporta y cómo falla.

Artículos relacionados

- IDE-019 — ¿Qué es la luz? Onda, cuanto y observador
- IDE-010 — Retroalimentación
- IDE-016 — Complejidad

Referencias

- [1] Duncan J. Watts; Steven H. Strogatz. Collective Dynamics of Small-World Networks. Nature 393, 440-442 (1998). DOI 10.1038/30918.
- [2] Albert-László Barabási; Réka Albert. Emergence of Scaling in Random Networks. Science 286, 509-512 (1999). DOI 10.1126/science.286.5439.509.
- [3] Réka Albert; Hawoong Jeong; Albert-László Barabási. Error and Attack Tolerance of Complex Networks. Nature 406, 378-382 (2000). DOI 10.1038/35019019.
- [4] Aaron Clauset; Cosma Rohilla Shalizi; M. E. J. Newman. Power-Law Distributions in Empirical Data. SIAM Review 51(4), 661-703 (2009). DOI 10.1137/070710111.
- [5] Mark S. Granovetter. The Strength of Weak Ties. American Journal of Sociology 78(6), 1360-1380 (1973). DOI 10.1086/225469.
- [6] M. E. J. Newman. Networks, 2nd edition. Oxford University Press (2018).

¿Qué es la luz? Onda, cuanto y observador

IDE-019 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.9 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: De las ondas de radio a los rayos gamma; de experimentos con fotones aislados a la astronomía

Límite: La luz en concreto: cómo una sola teoría produce fenómenos ondulatorios y cuánticos. Qué es una partícula en general corresponde a IDE-020, qué es un campo cuántico corresponde a IDE-021, y la observación como acto epistémico corresponde a IDE-023

Trabajos relacionados: IDE-020, IDE-021, IDE-023

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La luz no es onda los lunes y partícula los martes. Las ecuaciones de Maxwell la describen como una perturbación que se propaga en el campo electromagnético, y ese relato explica sin resto, dentro de su dominio, la interferencia, la difracción, la polarización y la velocidad finita de propagación.[1] El argumento de Einstein de 1905 añadió algo que el relato clásico no podía dar: la energía se intercambia con la materia en cantidades discretas proporcionales a la frecuencia, algo que Millikan midió después con cuidado suficiente para extraer la constante de Planck.[2,3] La reconciliación no es un compromiso entre dos imágenes sino una sola teoría en la que ambas familias de fenómenos son consecuencias. Este artículo examina también las pruebas no clásicas basadas en estadísticas de fotones, la fuerza probatoria limitada del efecto fotoeléctrico por sí solo y el papel físico de un observador.[4,7]

En 60 segundos

El electromagnetismo clásico acierta con la luz allí donde se aplica: la interferencia y la difracción se siguen de una perturbación de campo que obedece ecuaciones de onda, y la velocidad de propagación sale de dos constantes eléctricas medibles.[1] Lo que no puede hacer es explicar por qué la energía de los fotoelectrones depende de la frecuencia y no de la intensidad. Einstein propuso que la energía se intercambia en cuantos proporcionales a la frecuencia, y Millikan, que partió escéptico, midió la relación con precisión suficiente para determinar con ella la constante de Planck.[2,3] Ahora bien, la discreción del intercambio no es prueba de balas diminutas. El experimento que muestra que la luz no es una onda clásica con tiempos de llegada aleatorios es el antiagrupamiento, en el que un emisor único no puede producir dos detecciones a la vez, comportamiento sin descripción ondulatoria clásica.[4] El entrelazamiento inducido por el entorno suprime la interferencia entre registros macroscópicamente distintos.[5]

Por qué importa esta pregunta

La luz es donde casi todo el mundo se encuentra con la física cuántica, y es donde se instalan las peores formulaciones divulgativas. La dualidad onda-partícula sugiere dos sustancias que se turnan; el efecto del observador sugiere que mirar cambia la realidad; el fotón se imagina como una bolita luminosa con

trayectoria. Cada una de esas cosas dificulta aprender física después, porque hay que desaprenderla. Acertar con la luz fija además el vocabulario de los dos artículos siguientes, porque un fotón es el ejemplo más limpio disponible de qué puede y qué no puede significar la palabra partícula.

1. La luz como perturbación de un campo

Las ecuaciones de Maxwell admiten soluciones en las que los campos eléctrico y magnético se sostienen mutuamente y se propagan. La velocidad de esas soluciones queda fijada por dos constantes medibles en un laboratorio sin óptica alguna, y coincidía con la velocidad de la luz medida, razón por la que Maxwell concluyó que la luz es una perturbación electromagnética.[1]

El relato no es una metáfora. Predice cuantitativamente la interferencia y la difracción, explica la polarización como orientación del campo y da la presión de radiación. En su dominio es completo: ningún experimento de interferencia exige añadirle nada, y la teoría clásica sigue siendo la descripción de trabajo en radio, óptica y buena parte de la astronomía.

Lo que no describe es el intercambio de energía entre luz y materia a baja intensidad. Clásicamente, la energía entregada a un electrón debería crecer con el brillo de la iluminación, y una fuente tenue debería sencillamente tardar más en acumular lo que una brillante entrega deprisa. Esa predicción falla, y el fallo no es pequeño.

En qué punto estamos

La descripción de campo de Maxwell explica la interferencia, la difracción, la polarización, la presión de radiación y la velocidad de propagación, y sigue siendo la teoría de trabajo en casi todo el espectro. Su fallo es concreto y está localizado: se equivoca en el intercambio de energía con la materia. Ahí entra el cuanto, y entra como enunciado sobre el intercambio y no sobre la propagación.

2. El cuanto de intercambio

El artículo de Einstein de 1905 se titula como un punto de vista heurístico, y su propuesta es precisa: en los procesos de emisión y absorción, la luz se comporta como si su energía estuviera distribuida en cuantos localizados proporcionales a la frecuencia.[2] Aplicado al efecto fotoeléctrico predice que la energía cinética máxima de los electrones arrancados crece linealmente con la frecuencia, con pendiente igual a la constante de Planck, y no depende de la intensidad.

Millikan dedicó una década a ponerlo a prueba, y fue explícito en que esperaba que la hipótesis fallara. Sus medidas de 1916 confirmaron la relación lineal y dieron un valor de la constante de Planck a partir de la pendiente, que es una forma fuerte de confirmación: la predicción era cuantitativa y la constante obtenida concordaba con valores determinados por vías no relacionadas.[3]

Conviene notar qué establece el resultado y qué no. Establece que la energía se intercambia en cantidades discretas fijadas por la frecuencia. No establece que la luz viaje como un chorro de objetos pequeños, y que el propio Einstein lo presentara como heurístico refleja que el argumento trataba de la estadística del intercambio y no de una imagen de la propagación.

En qué punto estamos

Einstein propuso que la luz intercambia energía en cuantos proporcionales a la frecuencia, y Millikan midió la relación lineal resultante con precisión suficiente para extraer de su pendiente la constante de Planck. El resultado trata del intercambio, no de una imagen de qué viaja. Establecer que la luz no es en absoluto una onda clásica exigió otro experimento posterior.

3. Qué no es un fotón

El efecto fotoeléctrico admite un tratamiento semiclásico con un campo electromagnético clásico acoplado a materia cuantizada; por tanto, estadísticas fotoeléctricas importantes no requieren por sí solas cuantización del campo.[7] Inferencia: el efecto fotoeléctrico aislado constituye, en consecuencia, evidencia más débil de los fotones que una prueba óptica plenamente no clásica.

Kimble, Dagenais y Mandel observaron antiagrupamiento en la fluorescencia de un átomo único: tras una detección, la probabilidad de una segunda detección inmediata cae por debajo de lo que permite cualquier descripción ondulatoria clásica.[4] Un emisor único que acaba de emitir debe reexcitarse antes de volver a emitir, así que su salida no puede dar dos detecciones simultáneas, y ningún campo clásico, con cualquier perfil de fluctuación de intensidad, reproduce eso.

Esto autoriza la palabra fotón y a la vez la restringe. El antiagrupamiento en la fluorescencia de un emisor único viola un límite de campo clásico y respalda una descripción de campo cuantizado.[4,7] No muestra que un fotón tenga posición entre emisión y absorción, ni trayectoria, ni tamaño; y los experimentos de interferencia que la teoría clásica explica tan bien siguen siendo válidos, lo que significa que, sea lo que sea un fotón, no es un objeto pequeño que sigue un camino.

En qué punto estamos

El antiagrupamiento en la fluorescencia de un emisor único viola un límite de campo clásico y respalda una descripción de campo cuantizado.[4] Queda qué hace un detector cuando registra uno.

4. El observador es una interacción

La expresión efecto del observador invita a leer que un acto consciente cambia el resultado. Lo que ocurre físicamente es que un detector interactúa con el sistema y su estado queda correlacionado con lo que detectó, proceso físico que involucra muchos grados de libertad.

La decoherencia describe la consecuencia. El entrelazamiento inducido por el entorno suprime la interferencia entre registros macroscópicamente distintos.[5]

Lo que ese resultado cubre conviene enunciarlo con exactitud, porque es más estrecho que la versión divulgativa. La decoherencia explica la supresión de la interferencia y la emergencia de un conjunto preferente de estados estables, que es lo que hace que los registros macroscópicos se comporten clásicamente respecto de la interferencia; la pregunta de por qué se experimenta un resultado y no otro sigue abierta y se interpreta de manera distinta en las lecturas principales de la teoría cuántica.[5,8] El artículo hace la primera afirmación y declina la segunda.

En qué punto estamos

La decoherencia no zanja por qué se experimenta un resultado concreto, y ese límite es deliberado. Queda enunciar cómo una sola teoría entrega ambas familias de fenómenos.

5. Una teoría, dos familias de fenómenos

La electrodinámica cuántica es una sola teoría del campo electromagnético y de su interacción con la materia cargada, y tanto los fenómenos ondulatorios como los cuánticos son consecuencias tuyas. La presentación de Feynman deja la estructura inusualmente clara: se asignan amplitudes a las maneras alternativas en que podría ocurrir un resultado, y esas amplitudes se suman antes de tomar la probabilidad.[6]

La interferencia no es entonces un fenómeno ondulatorio que compite con una imagen corpuscular. Es lo que ocurre cuando varias alternativas aportan amplitudes al mismo suceso de detección, y por eso fotones enviados de uno en uno a través de dos rendijas construyen igualmente un patrón de interferencia a lo largo de muchos sucesos. La discreción tampoco compite: cada suceso es una detección, y el patrón está en la distribución de muchas.

La dualidad onda-partícula fue un andamio útil mientras se construía la teoría y es una mala descripción de la estructura terminada, porque sugiere dos cosas donde hay una. El resumen honesto es que la luz es una excitación del campo electromagnético, descrita por una teoría cuyas amplitudes interfieren y cuyas interacciones son discretas.

En qué punto estamos

La electrodinámica cuántica produce fenómenos ondulatorios y cuánticos desde una sola estructura, en la que las maneras alternativas de que ocurra un resultado aportan amplitudes que interfieren antes de tomar probabilidades. La dualidad era andamio, no resultado. Como un fotón resulta ser una excitación y no un objeto, IDE-020 pregunta qué puede seguir significando la palabra partícula.

Lo que sabemos y cómo lo sabemos

- El electromagnetismo clásico explica la interferencia, la difracción, la polarización, la presión de radiación y la velocidad de propagación de la luz, y sigue siendo la descripción de trabajo en su dominio.[1]
- La energía se intercambia entre luz y materia en cuantos proporcionales a la frecuencia; la relación fotoeléctrica es lineal en la frecuencia e independiente de la intensidad, y su pendiente da la constante de Planck.[2,3]
- El antiagrupamiento en la fluorescencia de un emisor único viola un límite de campo clásico y respalda una descripción de campo cuantizado.[4,7]

Lo que todavía no sabemos

- Por qué se registra un resultado concreto y no otro en un suceso individual, cuestión que la decoherencia no aborda y que las interpretaciones principales de la teoría cuántica responden de forma distinta.[5]
- Si puede definirse un observable de posición plenamente satisfactorio para un fotón, lo que sigue siendo una dificultad técnica y no una parte asentada de la teoría.
- Cómo describir el campo electromagnético en regímenes de curvatura fuerte del espaciotiempo, cuestión que pertenece a la relación no resuelta entre teoría cuántica de campos y gravitación.

Qué podría hacernos cambiar de opinión

- Una fuente reproducible de luz antiagrupada que pudiera describirse por completo con un campo estocástico clásico, medida correctamente la correlación de segundo orden a retardo cero por debajo de la cota clásica, eliminaría la prueba decisiva de que ninguna descripción de campo clásico da cuenta de la luz.[4]
- Una desviación medida de la dependencia lineal de la energía máxima de los fotoelectrones con la frecuencia, en condiciones donde los efectos de superficie y de carga espacial estén controlados como en las determinaciones clásicas, contradiría el cuanto de intercambio.[2,3]
- Un experimento que recuperara la interferencia entre registros de detector macroscópicamente distintos, tras el entrelazamiento con un entorno de muchos grados de libertad, contradiría el

relato de la decoherencia sobre por qué esa interferencia se vuelve inobservable bajo tal acoplamiento, que es lo que ese relato afirma; no afectaría a por qué se obtiene un resultado y no otro, cuestión que el relato no aborda.[5]

Si solo recuerdas una cosa

La luz es una excitación del campo electromagnético cuyas alternativas interfieren y cuyas interacciones son discretas; no es una onda y una partícula que se turnan.

Artículos relacionados

- IDE-020 — ¿Qué es una partícula?
- IDE-021 — ¿Qué es un campo?
- IDE-023 — ¿Qué es la observación?

Referencias

- [1] James Clerk Maxwell. A Dynamical Theory of the Electromagnetic Field. Philosophical Transactions of the Royal Society of London 155, 459-512 (1865). DOI 10.1098/rstl.1865.0008.
- [2] Albert Einstein. On a Heuristic Point of View Concerning the Production and Transformation of Light. Annalen der Physik 17, 132-148 (1905). DOI 10.1002/andp.19053220607.
- [3] Robert A. Millikan. A Direct Photoelectric Determination of Planck's h . Physical Review 7(3), 355-388 (1916). DOI 10.1103/PhysRev.7.355.
- [4] H. J. Kimble; M. Dagenais; L. Mandel. Photon Antibunching in Resonance Fluorescence. Physical Review Letters 39(11), 691-695 (1977). DOI 10.1103/PhysRevLett.39.691.
- [5] Wojciech H. Zurek. Decoherence, Einselection, and the Quantum Origins of the Classical. Reviews of Modern Physics 75(3), 715-775 (2003). DOI 10.1103/RevModPhys.75.715.
- [6] Richard P. Feynman. QED: The Strange Theory of Light and Matter. Princeton University Press (1985).
- [7] Leonard Mandel; Emil Wolf. Optical Coherence and Quantum Optics. Cambridge University Press (1995), chapter 9.
- [8] Guido Bacciagaluppi. The Role of Decoherence in Quantum Mechanics. The Stanford Encyclopedia of Philosophy, first published 3 November 2003, substantive revision 23 January 2025. <https://plato.stanford.edu/entries/qm-decoherence/>

EL UNIVERSO Y SUS OBSERVADORES

¿Qué es una partícula?

IDE-020 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.12 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Física subatómica, de las firmas en detectores a la teoría cuántica relativista

Límite: El propio concepto de partícula. Qué es un campo cuántico corresponde a IDE-021, la luz en concreto a IDE-019, y el significado de la masa a IDE-022

Trabajos relacionados: IDE-021, IDE-019, IDE-022

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La palabra partícula sobrevive en física casi sin nada de su contenido original. Una partícula clásica es un objeto pequeño con posición definida en cada instante que sigue una trayectoria. Los estados y las amplitudes cuánticos pueden interferir, y sus predicciones no quedan recogidas por una ontología de trayectorias de partículas clásicas. Lo que sustituye a la imagen clásica es una clasificación: Wigner mostró que los posibles estados de una partícula en una teoría cuántica relativista corresponden a representaciones irreducibles de la simetría del espaciotiempo, etiquetadas por invariantes como la masa y el espín.[1,2] Este artículo sigue qué conserva la palabra, incluido el resultado sorprendente de que el número de partículas no es una magnitud independiente del observador, y qué debe ceder. Lo que queda son sucesos discretos de detección, números cuánticos conservados y etiquetas invariantes, que es una respuesta de otro tipo que una imagen.[3,4]

En 60 segundos

Empieza por lo que sobrevive a la medida. Los detectores registran sucesos discretos con números cuánticos definidos, y esos son los hechos que un concepto de partícula debe acomodar.[5] Lo que no sobrevive es la imagen que los produce. En teoría cuántica de campos el número de partículas cambia: las interacciones crean y destruyen cuantos, así que no hay disponible un inventario fijo de objetos pequeños.[2] La identidad también cambia de carácter. Wigner clasificó los estados relativistas de una partícula por representaciones irreducibles del grupo de simetría del espaciotiempo, y por eso masa y espín son las etiquetas que persisten cuando falla la imagerie clásica.[1] El resultado más fuerte es que el número de partículas no es absoluto: un detector acelerado, en el estado que un observador distinto llama vacío, registra partículas, así que qué cuenta como partícula depende del estado de movimiento del observador.[4] La palabra sobrevive; el objeto pequeño que antes nombraba, no.

Por qué importa esta pregunta

La física de partículas se presenta al público en términos de bolitas, y esa imagerie es la fuente de casi toda la confusión posterior: sobre por qué las partículas de un tipo son idénticas, sobre cómo puede crearse una a partir de energía, sobre qué ocurre entre rendijas. Sustituir la imagen por lo que la teoría dice de verdad no es precisión por la precisión. Cambia qué preguntas tienen sentido, y es lo que vuelve inteligible y no paradójico el tema del artículo siguiente, el campo del que surgen las partículas.

1. La imagen clásica y dónde falla

Una partícula clásica es un objeto idealizado con posición, momento, masa y camino. La idealización tiene un éxito enorme para planetas y bolas de billar, y sobrevive en la física cuántica como aproximación siempre que las longitudes de onda pertinentes son pequeñas frente a todo lo demás.

Falla de maneras concretas y no vagas. Los estados y las amplitudes cuánticos pueden interferir, y sus predicciones no quedan recogidas por una ontología de trayectorias de partículas clásicas.[2] Esto no excluye formulaciones que añaden posiciones definidas de partículas y su movimiento a la descripción del estado: todo el formalismo cuántico, incluidos el principio de incertidumbre y la aleatoriedad cuántica, emerge de un análisis de ese movimiento.[7] Lo que no es opcional es que la trayectoria clásica, la de valores definidos y locales, no está disponible.

La respuesta correcta no es declarar que las partículas son ondas. Es preguntar qué sí proporciona la teoría: sucesos discretos de detección, números cuánticos conservados y estados clasificados por propiedades que sobreviven a los cambios de sistema de referencia. Es una respuesta de otro tipo que una imagen, y es lo que desarrolla el resto del artículo.

En qué punto estamos

Los estados cuánticos pueden estar extendidos e interferir, por lo que una ontología de trayectorias clásicas resulta inadecuada, aunque existen formulaciones que restauran posiciones definidas al precio de la no localidad.[2,7] Sustituir la imagen exige preguntar qué proporciona la teoría en su lugar. Lo primero que proporciona es una clasificación, y viene de la simetría y no de la imaginaria.

2. Partículas como representaciones de la simetría

Wigner preguntó qué estados cuánticos son posibles en una teoría compatible con la relatividad especial, y lo respondió clasificando las representaciones unitarias irreducibles del grupo de simetría del espaciotiempo.[1] La clasificación es exhaustiva, y las etiquetas que emergen son la masa y una magnitud de tipo espín: precisamente las propiedades con que se identifican las especies de partículas.

Es un resultado fuerte y conviene enunciarlo con cuidado. No dice que las partículas sean representaciones en el sentido de estar hechas de matemáticas. Dice que, sea lo que sea un estado de una partícula, sus propiedades de transformación bajo cambios de sistema inercial están fijadas, y que esas propiedades vienen en una lista discreta.[2] La masa y el espín son etiquetas invariantes porque son aquello con lo que se construye la clasificación.

Explica además un hecho que de otro modo resulta misterioso. Wigner clasifica las representaciones irreducibles del grupo de Poincaré mediante etiquetas invariantes.[1] Dos electrones son indistinguibles. No es una afirmación de que ocupen el mismo estado, cosa que en general no hacen: es una afirmación sobre las etiquetas. La teoría codifica la consecuencia admitiendo solo estados de varias partículas con un comportamiento definido al intercambiar dos de ellas. La identidad no es un logro de fabricación sino una consecuencia de qué es una partícula.

En qué punto estamos

Wigner clasifica las representaciones irreducibles del grupo de Poincaré mediante etiquetas invariantes. La clasificación trata de estados individuales. Qué le ocurre al recuento cuando las partículas interactúan es otro asunto y más extraño.

3. El número de partículas no es fijo

En una teoría cuántica relativista el número de partículas no se conserva. Las interacciones crean y aniquilan cuantos, y la descripción matemática natural usa operadores que los añaden y los quitan de un estado.[2] Un electrón y un positrón pueden aniquilarse en fotones; una colisión lo bastante energética produce partículas que antes no estaban.

Aquí la imagen clásica se vuelve insostenible y no meramente inexacta. Un objeto pequeño no deja de existir al encontrarse con otro objeto pequeño, pero una excitación cuántica sí. Lo que se conserva no es el recuento sino un conjunto de números cuánticos, y esos restringen qué procesos son posibles.[5]

Las entidades compuestas e inestables complican más la contabilidad. Un protón es un estado ligado cuya descripción interna no es una lista fija de constituyentes; una resonancia se identifica por un pico en una sección eficaz con una anchura relacionada con su vida media, y llamarla partícula es un enunciado sobre un polo de una amplitud de dispersión y no sobre un objeto que persiste.[5] La palabra cubre un espectro de casos, y su extremo más firme es la clasificación, no la imaginaria.

En qué punto estamos

El número de partículas no se conserva en teoría cuántica relativista: las interacciones crean y destruyen cuantos, y lo que persiste son números cuánticos conservados y no un recuento. Las resonancias inestables se identifican por rasgos de las amplitudes de dispersión y no por su permanencia. Pero incluso el recuento supone algo que la teoría no siempre proporciona.

4. El número de partículas depende del observador

Definir cuántas partículas contiene un estado exige separar los grados de libertad del campo en partes de frecuencia positiva y negativa, y esa separación depende de la noción de tiempo empleada. Para observadores inerciales en espaciotiempo plano todos coinciden, así que la cuestión no aparece en la física de laboratorio ordinaria.

Unruh mostró que sí aparece en general. Un detector que acelera a través del estado que los observadores inerciales llaman vacío detecta partículas en él.[4] Las dos descripciones no se contradicen: son respuestas distintas e igualmente válidas a una pregunta que presupone un sistema de referencia.

La consecuencia para el concepto es nítida. El número de partículas no es una propiedad invariante de un estado como sí lo es la carga. Se define relativo a una elección, y en espaciotiempos generales, incluidos los cosmológicos, no tiene por qué existir una elección privilegiada. La teoría cuántica algebraica de campos toma como base redes de álgebras de observables locales e introduce conceptos de partícula de forma asintótica cuando resultan aplicables.[3]

En qué punto estamos

Cuántas partículas contiene un estado depende de la separación de los modos del campo y por tanto del observador, y el resultado de Unruh muestra un detector acelerado detectando partículas en el vacío inercial. El número de partículas no es entonces invariante como lo es la carga. Queda decir qué sigue haciendo el concepto de forma fiable.

5. Qué sobrevive

Tres cosas sobreviven intactas. Los sucesos de detección son discretos: un detector cuenta o no cuenta, y una vez que el aparato se acopla a un entorno ese registro es estable.[6] Los números cuánticos se conservan y restringen lo que puede ocurrir. Y las etiquetas de Wigner siguen siendo las propiedades identificativas de una especie, y por eso una tabla de partículas es una tabla de masas, espines y cargas y no de formas y tamaños.[1,5]

Lo que se descarta es la imagen: la trayectoria, la posición definida entre sucesos, el inventario fijo y la idea de que las partículas son el mobiliario fundamental con el que se construyen los campos. La teoría cuántica algebraica de campos toma como base redes de álgebras de observables locales e introduce conceptos de partícula de forma asintótica cuando resultan aplicables.[3]

Esa inversión es el punto donde este artículo se detiene. Una partícula se describe mejor como una excitación discreta, clasificada por simetría, que un detector puede registrar, y la pregunta siguiente natural es de qué es excitación, que es el tema de IDE-021.

En qué punto estamos

Lo que sobrevive es la discreción en la detección, los números cuánticos conservados y las etiquetas invariantes de masa y espín; lo que se descarta es la trayectoria, el inventario fijo y la idea de que las partículas son la estructura primaria. IDE-021 recoge de qué es excitación esa excitación.

Lo que sabemos y cómo lo sabemos

- Wigner clasifica las representaciones irreducibles del grupo de Poincaré mediante etiquetas invariantes.[1,2]
- El número de partículas no se conserva en teoría cuántica relativista; las interacciones crean y destruyen cuantos mientras los números cuánticos conservados restringen qué procesos ocurren.[2,5]
- Un detector acelerado detecta partículas en el estado que los observadores inerciales llaman vacío, así que el número de partículas se define relativo a un observador y no es un invariante del estado.[4]
- Los sucesos de detección son discretos, y las especies se catalogan por masa, espín y carga y no por forma o tamaño.[1,5]

Lo que todavía no sabemos

- Si puede definirse un concepto satisfactorio de partícula en espaciotiempos curvos generales, donde no existe una separación privilegiada de los modos del campo.[3,4]
- Cómo caracterizar excitaciones de tipo partícula en regímenes de interacción fuerte donde no están disponibles los estados asintóticos que sirven para definirlas.
- Si la lista actual de especies refleja una estructura organizadora más profunda o es el subconjunto observado de una clasificación mayor.

Qué podría hacernos cambiar de opinión

- La detección reproducible de una especie cuyas propiedades de transformación bajo cambios de sistema inercial cayeran fuera de la clasificación de Wigner exigiría revisar el propio marco relativista.[1,2]

- Una medida directa que mostrara que un detector acelerado no registra partículas en el vacío inercial, controladas la aceleración y el acoplamiento del detector, contradiría la dependencia del número de partículas respecto del observador.[4]
- Evidencia experimental de que el número de partículas se conserva en un régimen donde el marco estándar predice creación, contabilizados todos los canales de desintegración y las ineficiencias del detector, invalidaría el relato de las interacciones en teoría de campos.[2,5]

Si solo recuerdas una cosa

Una partícula cuántica es una excitación discreta clasificada por simetría que un detector puede registrar, no un objeto pequeño con trayectoria, y hasta su número depende del observador.

Artículos relacionados

- IDE-021 — ¿Qué es un campo?
- IDE-019 — ¿Qué es la luz? Onda, cuanto y observador
- IDE-022 — ¿Qué significa la masa?

Referencias

- [1] Eugene P. Wigner. On Unitary Representations of the Inhomogeneous Lorentz Group. *Annals of Mathematics* 40(1), 149-204 (1939). DOI 10.2307/1968551.
- [2] Steven Weinberg. *The Quantum Theory of Fields, Volume I: Foundations*. Cambridge University Press (1995).
- [3] Rudolf Haag. *Local Quantum Physics: Fields, Particles, Algebras*, 2nd edition. Springer (1996).
- [4] W. G. Unruh. Notes on Black-Hole Evaporation. *Physical Review D* 14(4), 870-892 (1976). DOI 10.1103/PhysRevD.14.870.
- [5] F. Takahashi et al. (Particle Data Group). Review of Particle Physics. *International Journal of Modern Physics A* 41, 2630011 (2026). DOI 10.1142/S0217751X26300115.
- [6] Wojciech H. Zurek. Decoherence, Einselection, and the Quantum Origins of the Classical. *Reviews of Modern Physics* 75(3), 715-775 (2003). DOI 10.1103/RevModPhys.75.715.
- [7] D. Durr; S. Goldstein; N. Zanghi. Bohmian Mechanics as the Foundation of Quantum Mechanics. arXiv:quant-ph/9511016 (1995).

EL UNIVERSO Y SUS OBSERVADORES

¿Qué es un campo?

IDE-021 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguroola

Versión: v0.8 ES · 17 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Teoría cuántica relativista, de las pruebas de laboratorio al Modelo Estándar

Límite: El campo cuántico: qué es, qué son sus excitaciones y qué contiene su estado fundamental.

La idea organizadora general de campo corresponde a IDE-009, y el concepto de partícula a IDE-020

Trabajos relacionados: IDE-022, IDE-009, IDE-020

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

En la física fundamental actual los objetos básicos no son partículas sino campos, y los campos no son magnitudes clásicas extendidas por el espacio. Un campo cuántico asigna operadores, y no números, a regiones del espaciotiempo, y lo que un detector llama partícula es una excitación de uno de ellos.[1,2] Esa inversión explica cosas que la imagen corpuscular no puede: por qué todos los electrones son exactamente iguales, por qué el número de partículas cambia en las interacciones y por qué el estado fundamental no está vacío en el sentido ordinario. La teoría es además estrictamente local en un sentido técnico preciso que convive con las correlaciones que identificó Bell.[3] Este artículo expone la estructura, la evidencia y el punto en que la pregunta de qué es un campo cuántico deja de tener respuesta asentada.[5]

En 60 segundos

Un campo clásico pone un número en cada punto. Un campo cuántico pone allí un operador, así que los valores del campo no son magnitudes poseídas a la vez; lo que existe es un estado, y el campo es lo que actúa sobre él.[1] Las partículas son excitaciones de esos campos, y todo electrón lleva exactamente las mismas etiquetas de masa y carga: no son copias fabricadas sino excitaciones de una estructura subyacente, que difieren en los estados que ocupan y no en lo que los hace electrones.[1,6] Eso explica también por qué cambia el número de partículas, ya que los operadores que crean y destruyen excitaciones son los objetos naturales de la teoría. El estado fundamental tampoco es nada: tiene estructura con consecuencias medibles, y el ejemplo más limpio es la atracción entre dos placas conductoras sin carga.[4] La localidad sobrevive a todo esto en una forma precisa, porque las medidas en regiones separadas por intervalo espacial conmutan, así que ninguna señal controlable adelanta a la luz aunque las correlaciones que analizó Bell superen lo que permite cualquier modelo clásico local.[2,3]

Por qué importa esta pregunta

Casi todo relato divulgativo de la física fundamental se construye sobre partículas, y casi toda dificultad de esos relatos se disuelve al invertir el orden. Por qué son idénticos los electrones, cómo puede la energía volverse materia, de qué está hecho el vacío, cómo puede el entrelazamiento ser no local sin permitir señales: parecen paradojas desde el lado corpuscular y consecuencias desde el lado del campo.

Acertar con el orden disciplina además el lenguaje, porque deja claro qué enunciados son resultados de una teoría bien contrastada y cuáles son interpretaciones de ella.

1. De números a operadores

Un campo clásico es una asignación de valores a puntos que evoluciona bajo ecuaciones locales. Cuantizarlo sustituye esos valores por operadores: objetos matemáticos que actúan sobre estados en vez de poseer magnitudes definidas propias.[1] Preguntar por el valor de un campo cuántico en un punto es por tanto una pregunta equivocada del mismo modo que preguntar por la posición de un electrón entre medidas.

Lo que la teoría proporciona en su lugar es un estado, junto con operadores asociados a regiones del espaciotiempo cuyos valores esperados y correlaciones son lo que miden los experimentos. En la formulación algebraica esto se toma como primario: el contenido físico es una familia de álgebras de observables adscritas a regiones, y todo lo demás se deriva de cómo encajan esas álgebras.[2]

La construcción no es un adorno opcional. Los requisitos de simetría relativista, causalidad y positividad espectral imponen restricciones fuertes a la forma que puede tomar un campo cuántico, una vez especificados los supuestos adicionales de la construcción elegida, y el desarrollo de Weinberg deriva el formalismo de campos de esos requisitos en vez de postularlo.[1] El resultado es una restricción y no un teorema de unicidad: por eso la teoría de campos es el marco de trabajo, no una prueba de que ninguna teoría cuántica relativista pueda levantarse sobre otros supuestos.

En qué punto estamos

Cuantizar un campo sustituye valores en puntos por operadores que actúan sobre estados, así que no tiene valores poseídos a la vez, y el contenido físico está en los estados y en observables adscritos a regiones. Los requisitos de simetría relativista, causalidad y positividad espectral imponen restricciones fuertes a esa forma, una vez especificados los supuestos adicionales de la construcción elegida. Lo primero que explica son las partículas.

2. Las partículas como excitaciones

Cada especie de partícula corresponde a un campo, y una partícula es una excitación cuantizada de ese campo, en el mismo sentido en que un modo normal de un sistema vibrante transporta cuantos discretos de energía.[1] Los operadores que aparecen en la teoría añaden y quitan esas excitaciones, expresión formal de la creación y la aniquilación.

Las partículas de una especie dada son excitaciones del campo correspondiente, y comparten las etiquetas de masa, carga y espín que definen la especie, mientras que los estados que ocupan pueden diferir. Hay un solo campo del electrón, y las etiquetas que llevan sus excitaciones son propiedades de ese campo, que el Modelo Estándar trata como parámetros que lo caracterizan.[1,6] No hay que invocar ningún proceso de fabricación para explicar la identidad de las etiquetas, y allí donde dos electrones difieren, difieren en el estado que ocupa cada uno y no en lo que los hace electrones.

Fija además el estatuto de las transiciones de vacío a partículas que en divulgación suenan paradójicas. La energía no se convierte en una sustancia llamada materia; un estado con excitaciones de un campo pasa a ser un estado con excitaciones de otros, sujeto a las leyes de conservación. La contabilidad está en los números cuánticos, y es exacta.

En qué punto estamos

Una partícula es una excitación cuantizada de un campo. Todos los electrones llevan las mismas etiquetas de masa, carga y espín que definen la especie, mientras que los estados que ocupan pueden diferir, y la creación y la aniquilación son las operaciones naturales de la teoría. Las excitaciones presuponen un estado fundamental, y ese estado resulta no estar vacío.

3. El estado fundamental no es la nada

El vacío de una teoría cuántica de campos es el estado de energía más baja, y es un estado y no una ausencia. Tiene correlaciones entre regiones, fluctúa en el sentido de que los observables de campo tienen varianza no nula en él, y no es invariante bajo todas las transformaciones que ingenuamente cabría esperar.[1,2,5]

La evidencia más clara es mecánica. Casimir predijo que dos placas conductoras paralelas sin carga en el vacío se atraen, porque las condiciones de contorno que imponen cambian la estructura del campo entre ellas respecto del exterior.[4] La fuerza depende de la separación y no de ninguna carga del material, y se ha medido.

Conviene cuidado con lo que esto autoriza. Que el estado fundamental tenga estructura con consecuencias medibles está bien establecido, y el vacío es el estado fundamental del hamiltoniano del campo y no la nada ni el espacio vacío; formulaciones construidas desde lagrangianos clásicos distintos pueden resultar ser la misma teoría descrita de otra manera.[5] Esas son las afirmaciones que hace esta sección, y son más estrechas que la imagen divulgativa de un caldo hirviendo de partículas que aparecen y desaparecen. A lo que se mide y a lo que se imagina no debe asignárseles el mismo estatuto.

En qué punto estamos

El vacío es el estado de energía mínima y no una ausencia, con correlaciones y varianza no nula en los observables de campo, y la atracción de Casimir entre placas sin carga es evidencia medible de que las condiciones de contorno lo modifican. Cómo es el estado fundamental más allá de lo que esas medidas alcanzan no queda zanjado por ellas. Queda conciliar la localidad con las correlaciones que permiten los estados cuánticos.

4. Localidad, y qué mostró Bell

La teoría cuántica de campos relativista es local en un sentido preciso: los observables asociados a regiones que no pueden conectarse por una señal luminosa conmutan, lo que implica que una medida en una región no puede cambiar la estadística de ninguna medida en la otra.[2] Ninguna influencia controlable adelanta a la luz, y eso es un teorema del marco y no una esperanza.

Bell mostró algo distinto e igual de firme. Las correlaciones entre medidas sobre sistemas entrelazados pueden superar lo que permite cualquier modelo que atribuya valores locales preexistentes a los sistemas, y la cota que dedujo vuelve esto experimentalmente decidible.[3] La teoría cuántica predice violaciones de esa cota, y los experimentos las encuentran.

Ambos resultados conviven sin tensión porque tratan de cosas distintas. El análisis de Bell descarta una clase de explicación, los modelos locales de variables ocultas; la conmutación a separación espacial descarta una clase de uso, la señalización. Correlaciones más fuertes de lo que permite cualquier modelo clásico local no transmiten por sí mismas nada, porque ninguna de las partes puede dirigir la estadística marginal de la otra.[2,3]

En qué punto estamos

Los observables en regiones separadas por intervalo espacial conmutan, así que ninguna señal controlable adelanta a la luz, mientras que la cota de Bell muestra que las correlaciones superan lo que permiten valores locales preexistentes. Ambos restringen cosas distintas, la explicación y el uso, y por eso coexisten. Queda la pregunta que el formalismo no zanja: qué clase de cosa es un campo cuántico.

5. Qué clase de cosa es un campo cuántico

El éxito predictivo del marco no está en discusión. El momento magnético anómalo del electrón se mide con una precisión de alrededor de una parte en diez mil millones.[6] Esa medida se compara con un cálculo de electrodinámica cuántica, y lo bien que ambos concuerdan depende tanto de la incertidumbre del cálculo como de la de la medida, así que una cifra de concordancia no se lee de la barra de error experimental sola. Aun así, es un observable de la electrodinámica cuántica y no el marco entero, porque las predicciones del sector de interacción fuerte no se contrastan ni de lejos con esa precisión.

La pregunta interpretativa es otra y sigue abierta. Si las entidades fundamentales son campos, o partículas, o las álgebras locales de la formulación algebraica, o ninguna de esas cosas, se sigue discutiendo; formulaciones distintas de la misma física sostienen lecturas distintas, y la matemática no lo decide.[2,5] La defensa de Wallace de tomar en serio la formulación lagrangiana de trabajo como descripción, y no como recurso de cálculo a la espera de ser sustituido, es una posición en un debate vivo y no un consenso.[5]

Así que el artículo se detiene donde se detiene la evidencia. Los campos, en el sentido de las estructuras de operadores del Modelo Estándar, organizan todo lo que se ha medido sobre la materia elemental, y las partículas son sus excitaciones. Qué son en último término esas estructuras no lo responde ese éxito, y la versión honesta de este asunto lo dice.

En qué punto estamos

Un observable de la electrodinámica cuántica, el momento magnético anómalo del electrón, se mide con una precisión de una parte en diez mil millones, mientras que qué es en último término un campo cuántico sigue siendo una pregunta interpretativa abierta que la matemática no zanja. Un éxito predictivo de esa clase restringe descripciones sin seleccionar una. IDE-022 recoge una propiedad que llevan esas excitaciones y que se describe mal de forma rutinaria: la masa.

Lo que sabemos y cómo lo sabemos

- La cuantización sustituye los valores del campo en puntos por operadores que actúan sobre estados, y los requisitos de simetría relativista, causalidad y positividad espectral imponen restricciones fuertes a la forma que pueden tomar esos campos, una vez especificados los supuestos adicionales de la construcción elegida.[1]
- Las partículas son excitaciones cuantizadas de campos, todas las excitaciones de un campo dado llevan idénticas etiquetas de masa, carga y espín mientras que los estados que ocupan pueden diferir, y el número de partículas cambia en las interacciones.[1,6]
- El vacío es el estado de energía más baja y no una ausencia, y tiene estructura con consecuencias medibles, de las que la atracción de Casimir entre placas conductoras sin carga es la más clara.[1,4,5]

- Los observables en regiones separadas por intervalo espacial conmutan, así que ninguna señal controlable adelanta a la luz, mientras que la cota de Bell muestra correlaciones que superan lo que permiten valores locales preexistentes.[2,3]

Lo que todavía no sabemos

- Cuáles son las entidades fundamentales de la teoría, con campos, partículas y álgebras locales defendidos y un formalismo que no determina la elección.[2,5]
- Cómo formular con pleno rigor matemático teorías cuánticas de campos con interacción en cuatro dimensiones espaciotemporales, cuestión abierta pese al éxito predictivo del marco.
- Cómo combinar la descripción de campos con la gravitación, que es donde se espera que el marco necesite sustitución y no extensión.

Qué podría hacernos cambiar de opinión

- Una demostración reproducible de señalización controlable entre regiones separadas por intervalo espacial contradiría la estructura de conmutación sobre la que se construye la teoría cuántica de campos relativista.[2]
- Una fuerza de Casimir medida que difiriera de la predicción para una geometría y conductividad bien caracterizadas, más allá de las correcciones conocidas por conductividad finita, rugosidad y temperatura, socavaría el relato estructural del vacío.[4]
- Una discrepancia confirmada entre el momento magnético anómalo del electrón calculado y medido, mayor que las incertidumbres teórica y experimental combinadas, indicaría que el marco es incompleto a energías accesibles.[6]

Si solo recuerdas una cosa

Un campo cuántico es una estructura de operadores cuyas excitaciones son partículas, todos los electrones llevan las mismas etiquetas de especie, y el vacío es un estado, no una ausencia.

Artículos relacionados

- IDE-022 — ¿Qué significa la masa?
- IDE-009 — Campo
- IDE-020 — ¿Qué es una partícula?

Referencias

- [1] Steven Weinberg. The Quantum Theory of Fields, Volume I: Foundations. Cambridge University Press (1995).
- [2] Rudolf Haag. Local Quantum Physics: Fields, Particles, Algebras, 2nd edition. Springer (1996).
- [3] John S. Bell. On the Einstein Podolsky Rosen Paradox. Physics Physique Fizika 1(3), 195-200 (1964). DOI 10.1103/PhysicsPhysiqueFizika.1.195.
- [4] H. B. G. Casimir. On the Attraction Between Two Perfectly Conducting Plates. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen 51, 793-795 (1948).
- [5] David Wallace. In Defence of Naiveté: The Conceptual Status of Lagrangian Quantum Field Theory. Synthese 151(1), 33-80 (2006). DOI 10.1007/s11229-004-6248-9.
- [6] F. Takahashi et al. (Particle Data Group). Review of Particle Physics. International Journal of Modern Physics A 41, 2630011 (2026). DOI 10.1142/S0217751X26300115.

¿Qué significa la masa?

IDE-022 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguroola

Versión: v0.7 ES · 17 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: De las partículas elementales al comportamiento gravitatorio de los cuerpos

Límite: La masa como propiedad física: qué es, de dónde viene y qué dice y qué no dice $E = mc^2$ al cuadrado. Qué es una partícula corresponde a IDE-020 y qué es un campo a IDE-021

Trabajos relacionados: IDE-023, IDE-020, IDE-021

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

La masa no es la cantidad de materia de un objeto, y la masa de la materia ordinaria no se la da principalmente el campo de Higgs. La masa invariante es una propiedad de todo un sistema, definida a partir de su energía y su momento totales, y no es aditiva: un sistema ligado suele tener menos masa que sus partes separadas, y un cuerpo caliente más que uno frío.[1] El protón ilustra ambas cosas. Sus quarks aportan solo una fracción pequeña de su masa; el resto viene de la energía de la interacción fuerte que los liga, como confirman hoy los cálculos en red desde primeros principios.[4] Este artículo separa la masa invariante del uso obsoleto de masa relativista, enuncia de qué da cuenta realmente el mecanismo de Higgs y cierra con la igualdad contrastada entre masa inercial y gravitatoria.[3,5]

En 60 segundos

En relatividad especial la energía y el momento de un sistema dependen del sistema de referencia, y la combinación que no depende es su masa invariante. Okun sostuvo que esa única magnitud debe llevar el nombre y que la masa dependiente de la velocidad debe abandonarse, porque conservarla invita a la imagen falsa de un objeto que se vuelve más masivo al acelerarse cuando lo que cambia es su energía.[1] El resultado de Einstein de 1905 se lee entonces bien: un cuerpo que emite energía pierde masa, así que la masa mide contenido energético y no es una sustancia convertible en energía.[2] La materia ordinaria lo muestra con nitidez. La cromodinámica cuántica en red calcula masas de hadrones desde primeros principios y encuentra que la mayor parte de la masa del protón es la energía del campo gluónico y el movimiento de sus quarks, no las masas en reposo de los quarks.[4] El mecanismo de Higgs da cuenta de las masas de los bosones gauge débiles y, mediante acoplamientos aparte, de los fermiones elementales, que es una porción mucho menor de la masa que te rodea.[3,6]

Por qué importa esta pregunta

Dos afirmaciones sobre la masa se repiten en todas partes y ambas son falsas de maneras que importan. La primera es que $E = mc^2$ al cuadrado describe materia que se convierte en energía, lo que confunde una identidad con un proceso de conversión. La segunda es que el campo de Higgs da a las cosas su masa, cierto para las partículas elementales y falso para los átomos que componen a una persona, donde la inmensa mayoría de la masa es energía de ligadura. Corregirlas no es pedantería: la primera gobierna cómo se explica la energía nuclear, y la segunda decide si el descubrimiento del Higgs se entiende.

1. La masa invariante pertenece a un sistema

En relatividad especial la energía y el momento dependen del sistema de referencia, y la combinación invariante de ambos es la masa del sistema. Es la misma para todo observador inercial, que es lo que la convierte en propiedad del sistema y no de una descripción suya.[1]

Tampoco es aditiva. Dos fotones, cada uno sin masa, forman un sistema con masa invariante no nula si viajan en direcciones distintas, porque los momentos se cancelan en parte mientras las energías se suman. Un núcleo ligado pesa menos que sus nucleones separados en la energía de ligadura dividida por la velocidad de la luz al cuadrado. Un ladrillo caliente tiene algo más de masa que el mismo ladrillo frío.[1]

El argumento de Okun era que la palabra debe reservarse para ese invariante y que la masa relativista, magnitud que crece con la velocidad, debe abandonarse. Conservarla produce la impresión de que la velocidad vuelve más masivo un objeto en algún sentido intrínseco, cuando lo que cambia es su energía en un sistema elegido y el invariante queda intacto.[1]

En qué punto estamos

La masa invariante es la combinación independiente del sistema de referencia de la energía y el momento, y no es aditiva: la ligadura la baja y la energía interna la sube. Reservar la palabra para el invariante evita la impresión de que la velocidad hace más pesadas las cosas. Ese encuadre resuelve además qué afirma realmente la ecuación más citada de la física.

2. Qué dice $E = mc^2$

El breve artículo de Einstein de 1905 pregunta si la inercia de un cuerpo depende de su contenido energético y concluye que un cuerpo que emite radiación pierde masa en proporción a la energía emitida.[2] La relación se enuncia como una equivalencia entre masa y contenido energético, no como una receta para convertir una sustancia en otra.

Leídos así, los casos habituales se vuelven directos. Un reactor nuclear no aniquila materia: convierte un sistema con más energía de ligadura en productos con menos, y la diferencia de masa aparece como energía cinética y radiación. Nada se destruyó; la contabilidad se reorganizó, y el defecto de masa mide cuánto.

La ecuación tiene además un alcance que se exagera con regularidad. En la forma que suele citarse da la energía en reposo de un sistema, y la relación general incluye el momento. Un fotón tiene masa nula, medida como compatible con cero por debajo de 1×10^{-18} eV,[6] y energía no nula, cosa imposible si la forma citada se toma como definición universal de la energía.[1] La relación completa cubre ambos casos; la abreviada no.

En qué punto estamos

El resultado de Einstein dice que un cuerpo que emite energía pierde masa, lo que hace de la masa una medida del contenido energético y no una sustancia convertible en energía. La forma citada da solo la energía en reposo, y las partículas sin masa que transportan energía muestran por qué hace falta la relación general. Queda dónde reside realmente la masa de la materia cotidiana.

3. De dónde viene la masa ordinaria

Casi toda la masa de la materia ordinaria está en los núcleos atómicos, y casi toda esa está en protones y neutrones. Así que la pregunta de qué compone tu masa es en gran medida la pregunta de qué compone la de un protón.

La respuesta no son sus quarks. Las masas de los quarks arriba y abajo son minúsculas frente al protón, y la inmensa mayoría de su masa viene de la energía de la interacción fuerte: el campo gluónico y la energía cinética de los quarks confinados.[4] Es el principio de la sección anterior en funcionamiento, porque esa energía contribuye a la masa invariante del sistema ligado exactamente como exige la relación.

La afirmación está calculada y no simplemente enunciada. Dürr y colaboradores calcularon masas de hadrones ligeros desde la teoría subyacente con métodos de red, con las masas de los quarks y una escala global como únicas entradas, y reprodujeron el espectro observado.[4] Es una confirmación desde primeros principios de que la masa de la materia ordinaria es predominantemente energía de interacción.

En qué punto estamos

La mayor parte de la masa de la materia ordinaria es la energía de la interacción fuerte dentro de protones y neutrones, y no las masas en reposo de los quarks, y los cálculos en red reproducen el espectro hadrónico desde la teoría para demostrarlo. Ese resultado fija además el alcance del mecanismo al que más a menudo se atribuye dar masa a las cosas.

4. De qué da cuenta el mecanismo de Higgs

Higgs mostró cómo los bosones gauge pueden adquirir masa sin estropear la simetría gauge de la teoría, mediante un campo con valor esperado en el vacío no nulo.[3] El fenómeno se llama convencionalmente ruptura espontánea de la simetría gauge, y el nombre debe leerse con cuidado. Una simetría gauge local no se rompe espontáneamente en el sentido invariante gauge, y lo que puede romperse una vez fijado el gauge es un subgrupo global remanente de ella.[7] Leer ese resultado como prueba de que la simetría es una redundancia de la descripción y no un rasgo del estado físico es la lectura propia de este artículo, compartida con IDE-005. El contenido invariante gauge es que los bosones gauge se vuelven masivos. En el Modelo Estándar eso da cuenta de las masas de los bosones W y Z, y el mismo campo da masa a los fermiones elementales mediante acoplamientos aparte cuyas intensidades son parámetros libres ajustados a la medida.[6]

De lo que no da cuenta es de casi toda la masa de la habitación. La masa del electrón sí procede de su acoplamiento al campo de Higgs, y también las pequeñas masas de los quarks, pero esas masas aportan solo una fracción pequeña de la de un protón.[4,6] El mecanismo explica por qué las partículas elementales tienen masa siquiera en una teoría que de otro modo lo prohibiría, logro distinto y más específico de lo que sugiere el resumen divulgativo.

Merece la pena conservar la distinción porque sitúa el contenido real del descubrimiento. Hallar el bosón de Higgs confirmó un mecanismo concreto para un problema concreto del Modelo Estándar. No explicó de dónde viene tu peso, y decir lo contrario dificulta enunciar el resultado verdadero.

En qué punto estamos

El mecanismo de Higgs da cuenta de las masas de los bosones gauge débiles y, mediante acoplamientos aparte, de los fermiones elementales, y por eso hay partículas con masa en una teoría que de otro modo lo prohibiría. No da cuenta de casi nada de la masa de la materia ordinaria. Queda una propiedad de la masa aún sin tocar: que la misma magnitud gobierna la inercia y la gravedad.

5. Una magnitud, dos papeles

La masa aparece dos veces en la física clásica: como resistencia a la aceleración y como fuente y receptora de la gravitación. Nada en la mecánica newtoniana exige que sean el mismo número, y sin embargo el experimento las encuentra iguales con precisión extraordinaria.[5]

Esa igualdad es el principio de equivalencia débil, también llamado universalidad de la caída libre, y es el primer escalón de una jerarquía y no toda ella.[5] Añadir que el resultado de cualquier experimento local no gravitatorio es independiente de la velocidad y del lugar del laboratorio en caída libre da el principio de equivalencia de Einstein, que es lo que autoriza la descripción geométrica de la gravitación; extender el mismo enunciado a cuerpos autogravitantes y a experimentos gravitatorios da la forma fuerte, que la relatividad general satisface y la mayoría de las alternativas no. Los contrastes a lo largo de esa jerarquía están entre los experimentos más importantes de la física, y continúan porque una violación en cualquier nivel sería evidencia de física nueva.

Con esto el artículo cierra donde empezó. La masa no mide cuánta materia hay presente; es un único invariante que cuantifica el contenido energético de un sistema en su sistema en reposo, procede sobre todo de energía de interacción en la materia ordinaria, y gobierna a la vez la inercia y la gravitación. Cada una de esas cosas es una afirmación medida y no una definición.

En qué punto estamos

La masa inercial y la gravitatoria son iguales con precisión experimental extraordinaria, que es el principio de equivalencia débil y el primer escalón de una jerarquía cuya forma de Einstein autoriza la teoría geométrica, así que los contrastes continúan en todos los niveles. La masa es entonces un invariante que juega dos papeles y no una medida de cantidad de materia. IDE-023 pasa de lo que se mide a qué es medir.

Lo que sabemos y cómo lo sabemos

- La masa invariante es la combinación independiente del sistema de referencia de la energía y el momento de un sistema y no es aditiva: la energía de ligadura la reduce y la energía interna la aumenta.[1]
- El resultado de Einstein de 1905 establece que un cuerpo que emite energía pierde masa, así que la masa mide contenido energético en vez de nombrar una sustancia que se convierte en energía.[2]
- La mayor parte de la masa de un protón procede de la energía de la interacción fuerte y no de las masas en reposo de los quarks, y los cálculos en red reproducen el espectro de hadrones ligeros desde la teoría con las masas de los quarks y una escala como entradas.[4]
- El mecanismo de Higgs da cuenta de las masas de los bosones gauge débiles y, mediante acoplamientos aparte, de los fermiones elementales;[3,6] una simetría gauge local no se rompe en el sentido invariante gauge, y lo que puede romperse una vez fijado el gauge es un subgrupo global remanente;[7] la masa inercial y la gravitatoria son experimentalmente iguales con precisión muy alta, que es el principio de equivalencia débil.[5]

Lo que todavía no sabemos

- Por qué los acoplamientos de los fermiones al campo de Higgs, y con ellos las masas de las partículas elementales, toman los valores que toman, cosa que el Modelo Estándar parametriza en vez de explicar.[6]
- Si las masas de los neutrinos surgen por el mismo mecanismo que las de los demás fermiones o requieren estructura adicional, cuestión que sigue abierta.[6]
- Si la igualdad entre masa inercial y gravitatoria se cumple de forma exacta, algo que los experimentos siguen contrastando con precisión creciente sin garantía teórica.[5]

Qué podría hacernos cambiar de opinión

- Una violación medida de la igualdad entre masa inercial y gravitatoria, por encima de la sensibilidad de los contrastes actuales y reproducida de forma independiente, falsaría el principio de equivalencia débil y, con él, el cimiento de la teoría geométrica.[5]
- Un cálculo en red desde la teoría establecida que no reprodujera el espectro hadrónico observado, con errores sistemáticos y de discretización controlados, socavaría el relato de la masa ordinaria como energía de interacción.[4]
- Un caso demostrado en el que la masa invariante de un sistema aislado cambiara con el sistema de referencia del observador, correctamente medida, contradiría la definición de masa usada en toda la física relativista.[1]

Si solo recuerdas una cosa

La masa es el contenido energético invariante de un sistema, no su cantidad de materia, y casi toda la tuya es energía de la interacción fuerte y no algo que aportara el campo de Higgs.

Artículos relacionados

- IDE-023 — ¿Qué es la observación?
- IDE-020 — ¿Qué es una partícula?
- IDE-021 — ¿Qué es un campo?

Referencias

- [1] Lev B. Okun. The Concept of Mass. *Physics Today* 42(6), 31-36 (1989). DOI 10.1063/1.881171.
- [2] Albert Einstein. Does the Inertia of a Body Depend Upon Its Energy-Content? *Annalen der Physik* 18, 639-641 (1905). DOI 10.1002/andp.19053231314.
- [3] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters* 13(16), 508-509 (1964). DOI 10.1103/PhysRevLett.13.508.
- [4] S. Dürr; Z. Fodor; J. Frison; C. Hoelbling; R. Hoffmann; S. D. Katz; S. Krieg; T. Kurth; L. Lellouch; T. Lippert; K. K. Szabó; G. Vulvert. Ab Initio Determination of Light Hadron Masses. *Science* 322, 1224-1227 (2008). DOI 10.1126/science.1163233.
- [5] Clifford M. Will. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity* 17, 4 (2014). DOI 10.12942/lrr-2014-4.
- [6] F. Takahashi et al. (Particle Data Group). Review of Particle Physics. *International Journal of Modern Physics A* 41, 2630011 (2026). DOI 10.1142/S0217751X26300115.
- [7] Simon Friederich. Gauge symmetry breaking in gauge theories, in search of clarification. *European Journal for Philosophy of Science* 3(2), 157-182 (2013). DOI 10.1007/s13194-012-0061-y. Preprint arXiv:1107.4664v2.

¿Qué es la observación?

IDE-023 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.20 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Del ojo desnudo a detectores que registran lo que ningún sentido alcanza

Límite: La observación como acto epistémico y físico: qué cuenta como observar, qué aporta la teoría y en qué consiste el problema cuántico de la medida. La maquinaria operativa de la medición, del mensurando al presupuesto de incertidumbre, corresponde a IDE-014

Trabajos relacionados: IDE-024, IDE-014, IDE-019

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Observar no es mirar. Es una interacción que deja un registro, junto con una inferencia desde ese registro hasta una afirmación sobre el mundo, y la inferencia siempre usa teoría. La tesis de Hanson de que lo que uno ve depende de lo que sabe suele tomarse como la afirmación de que la evidencia es arbitraria; no lo es, y la respuesta de Hacking, que creemos en aquello que podemos manipular para producir efectos nuevos, muestra por qué.[1,2] La teoría cuántica añade un problema genuinamente distinto, que no es error instrumental y no se reduce con mejor calibración: Bell sostuvo que la palabra medida debía desterrarse de la formulación porque disfraza dónde está la dificultad.[3] Este artículo separa las componentes epistémica, física y sociológica que la palabra observación confunde.

En 60 segundos

Una observación empieza con una interacción física: algo se acopla a un aparato o a un entorno y deja un registro, sea un recuento, una tensión, una traza o una marca. Este artículo considera ese registro evidencia solo una vez interpretado mediante modelos del instrumento y del proceso, que es lo que hace la observación cargada de teoría en el sentido modesto y correcto.[1] La carga teórica no vuelve arbitraria la evidencia: las técnicas independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método, y los instrumentos pueden utilizarse para intervenir y no solo para observar.[2] La teoría cuántica plantea una dificultad aparte. No es que los instrumentos perturben lo que miden, cosa cierta también en el caso clásico, sino que la teoría tal como se enuncia divide el mundo en un sistema cuántico y un aparato clásico sin decir dónde cae la línea; Bell lo llamó una ambigüedad fundamental y no técnica.[3] La decoherencia explica por qué el entorno destruye la interferencia entre los estados que sobreviven a él, que es por lo que el mundo resulta efectivamente clásico, y no resuelve por qué se obtiene un resultado y no otro.[4]

Por qué importa esta pregunta

La palabra observación cubre tres cosas distintas que se confunden sin parar. Hay un proceso físico, que es lo que hace un detector. Hay una inferencia, donde entra la teoría y donde viven los errores interesantes. Y hay un proceso social por el que una comunidad llega a aceptar que algo se observó, donde trabajan la replicación y la calibración. Las discusiones divulgativas sobre si la observación crea

la realidad, o si los científicos solo ven lo que esperan, mezclan las tres; separarlas vuelve tratables ambas preguntas.

1. Una interacción que deja un registro

Físicamente, observar es acoplar. Se absorbe un fotón, se deposita una carga, se une una molécula, se revela un grano, y después alguna parte del aparato queda en un estado correlacionado con el sistema. Esa correlación es el registro, y es lo que pone el suceso a disposición de la inferencia.

Nada en esa descripción exige una persona. Una placa fotográfica expuesta en una habitación cerrada ha observado en todos los sentidos que importan físicamente, y el registro espera a que alguien lo lea. El papel del científico empieza en la lectura y no en la interacción, y por eso las apelaciones a la consciencia en la física de la medida están fuera de lugar.[4]

La evidencia, en el sentido que este artículo da a la palabra, es un registro ya interpretado. Una tensión se vuelve temperatura mediante una calibración y un modelo del sensor; una traza se vuelve identificación de partícula mediante un modelo del detector y de cómo se comportan tales partículas en él. La cadena del registro a la afirmación es donde está el trabajo, y IDE-014 trata su forma metrológica.[6]

En qué punto estamos

La observación es físicamente un acoplamiento que deja un registro correlacionado con el sistema, y no requiere observador en sentido mental. Este artículo llama evidencia a ese registro solo una vez aplicados los modelos del instrumento y del proceso. Esa dependencia de la teoría plantea inevitablemente la pregunta siguiente: si deja la evidencia rehén de la expectativa.

2. La carga teórica, bien enunciada

Hanson sostuvo que observar no es registrar neutralmente datos sensoriales: lo que un observador cree ver depende en parte de los conceptos y expectativas que trae, de modo que observadores científicos con formación pueden informar de que ven cosas distintas ante la misma preparación observada.[1] La tesis trata de la organización de la percepción y está bien sostenida.

Se sobreinterpreta con regularidad como la tesis de que la evidencia no puede arbitrar entre teorías, y eso no se sigue. Las técnicas experimentales independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método.[2]

La aportación de Hacking fue desplazar el terreno de ver a hacer. Lo que nos convence de que una entidad es real, argumentó, no es que una teoría lo diga sino que podemos manipularla para producir fenómenos nuevos: si puedes rociar electrones sobre algo y observar un efecto diseñado, la entidad se ha vuelto herramienta y no hipótesis.[2] La intervención da a la observación un control que el mirar pasivo no puede dar.

En qué punto estamos

Lo que un observador cree ver depende de los conceptos que aporta, algo bien respaldado; las técnicas independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método, mientras que la intervención comprueba lo que la mera observación no puede.[1,2] La teoría cuántica, sin embargo, plantea una dificultad que ninguna comprobación cruzada elimina.

3. El problema de la medida no es error instrumental

Es frecuente oír que en física cuántica la observación perturba el sistema. La perturbación es real, y también es enteramente familiar en el caso clásico: un termómetro cambia la temperatura de lo que mide. Si eso fuera todo, no habría problema sino una corrección que aplicar.[6]

La objeción de Bell es más aguda. La formulación estándar de la mecánica cuántica da una regla para la evolución no perturbada y otra para lo que ocurre en la medida, y no dice qué cuenta físicamente como medida ni dónde se traza la frontera entre sistema cuántico y aparato clásico.[3] Como el aparato está hecho de átomos, la teoría tal como se enuncia aplica sus dos reglas sin especificar cuándo cambiar de una a otra, y eso es una ambigüedad en los fundamentos y no un inconveniente práctico.

Conviene mantenerlo separado de la incertidumbre metrológica, que es un parámetro que caracteriza la dispersión de los valores atribuidos a un mensurando a partir de la información utilizada.[6] Mejorar la calibración reduce lo segundo y no hace nada con lo primero, y confundirlos vuelve invisible el problema real.

En qué punto estamos

El problema cuántico de la medida no es que los instrumentos perturben los sistemas, cosa clásicamente familiar, sino que la formulación estándar aplica dos reglas distintas sin decir dónde está la frontera entre sistema y aparato. Es una ambigüedad de fundamentos y no una cuestión de calibración. La decoherencia es la respuesta física más sustancial, y conviene decir qué entrega.

4. Qué zanja la decoherencia

Cuando un sistema interactúa con un entorno de muchos grados de libertad, el entorno vigila ciertos de sus observables, y la información que distingue sus alternativas se pierde selectivamente en correlaciones con ese entorno. La interferencia entre los estados que corresponden a esos observables queda por tanto destruida, y queda seleccionado un conjunto particular de estados, los que permanecen estables bajo esa interacción.[4]

Es un resultado físico real. Explica por qué queda destruida la interferencia entre esos estados y por qué los estados que sobreviven a la interacción son los que son. El mundo cotidiano queda así efectivamente clásico: el proceso descarta la inmensa mayoría de los estados que el formalismo permite, y en especial las superposiciones de estados macroscópicamente distintos.

Lo que no hace es explicar por qué ocurre un resultado y no otro. El formalismo da un conjunto de alternativas con probabilidades y una explicación de por qué ya no interfieren; el paso a un único resultado realizado es lo que las interpretaciones principales responden de forma distinta, y ese límite está enunciado explícitamente en la literatura de revisión sobre decoherencia.[4,7] Presentar la decoherencia como solución del problema entero exagera un buen resultado.

En qué punto estamos

La decoherencia explica por qué el entorno destruye la interferencia entre los estados que sobreviven y por qué se seleccionan esos estados concretos, pero no por qué ocurre un resultado y no otro. Presentarla como la solución completa la exagera. Más allá de la física queda la cuestión de cómo decide una comunidad que hubo observación.

5. Cómo decide una comunidad qué se observó

Decidir si un resultado es real suele significar replicarlo, y Collins señaló una dificultad en ese procedimiento. Para saber si un intento de replicación fue competente hace falta saber si el efecto es real; para saber si el efecto es real hacen falta replicaciones competentes. Lo llamó la regresión del experimentador, y lo documentó en casos donde la existencia del fenómeno estaba genuinamente en disputa.[5]

La regresión es real. En el caso documentado por Collins, la calibración frente a un efecto conocido es una práctica comunitaria utilizada para establecer la competencia experimental, aunque un discrepante puede rechazar el efecto de calibración como sustituto del efecto discutido.[5] Otras dos prácticas de laboratorio inciden en la misma dificultad sin determinar cómo la resuelven las comunidades. Las técnicas independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método; construir un aparato que produzca el fenómeno es lo que Hacking denomina creación de fenómenos.[2]

En conjunto, los tres componentes ofrecen una explicación operativa: la observación es una interacción física que produce un registro; los modelos convierten ese registro en evidencia y pueden comprobarse mediante técnicas independientes; y, cuando se discute la competencia, las prácticas comunitarias incluyen la calibración, que un discrepante puede rechazar.[2,5]

En qué punto estamos

Juzgar si una replicación fue competente puede exigir saber si el efecto es real. En el caso documentado por Collins, la calibración es una práctica comunitaria utilizada para establecer la competencia, y una calibración puede rechazarse.[5] Las técnicas independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método.[2] La observación es física, inferencial y social. IDE-024 recoge qué debe estar presente: una variación distinguible del fondo.

Lo que sabemos y cómo lo sabemos

- La observación es físicamente una interacción que deja un registro correlacionado con el sistema y no requiere observador en ningún sentido mental.[3,4]
- Lo que un observador cree ver depende de los conceptos que aporta; las técnicas independientes y las comprobaciones cruzadas pueden ayudar a distinguir los fenómenos de los artefactos específicos del método, y la intervención aporta una comprobación adicional.[1,2]
- El problema cuántico de la medida consiste en que la formulación estándar aplica dos reglas dinámicas distintas sin especificar la frontera entre sistema y aparato, cosa distinta de la perturbación instrumental y de la incertidumbre metrológica.[3,6]
- La decoherencia explica la destrucción de la interferencia entre los estados que sobreviven a la interacción y la selección de esos estados preferentes, pero no por qué ocurre un resultado concreto.[4]

Lo que todavía no sabemos

- Por qué se realiza un resultado concreto en un suceso cuántico individual, cuestión que las interpretaciones principales responden de forma distinta.[3,4]
- Dónde debería trazarse, si en algún sitio, una frontera fundamentada entre descripción cuántica y clásica, que Bell identificó como no especificada en la formulación estándar.[3]
- Cuánto de la resolución práctica de las observaciones disputadas depende de la evidencia y cuánto de factores extraevidenciales, cosa discutida en el estudio de la práctica científica.[5]

Qué podría hacernos cambiar de opinión

- Una demostración experimental que restituyera la interferencia entre registros de detector macroscópicamente distintos, tras el entrelazamiento con un entorno de muchos grados de libertad, contradiría el relato de la decoherencia sobre por qué esa interferencia queda destruida bajo tal acoplamiento, que es lo que ese relato afirma; no afectaría a por qué se obtiene un resultado y no otro, cuestión que el relato no aborda.[4]
- Una formulación de la mecánica cuántica que especificara una frontera física entre sistema y aparato e hiciera predicciones contrastables distintas de la teoría estándar resolvería la ambigüedad que identificó Bell, y podría comprobarse.[3]
- Una clase documentada de casos en que instrumentos independientes con errores sistemáticos genuinamente distintos convergieran en resultados demostrados después artefactuales debilitaría la coincidencia entre instrumentos como evidencia contra un artefacto.[1,2]

Si solo recuerdas una cosa

Observar es una interacción que deja un registro, una inferencia que usa teoría y un juicio comunitario de que puede repetirse; confundir las tres cosas lo vuelve misterioso.

Artículos relacionados

- IDE-024 — ¿Qué es una señal?
- IDE-014 — Medición
- IDE-019 — ¿Qué es la luz? Onda, cuanto y observador

Referencias

- [1] Norwood Russell Hanson. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge University Press (1958).
- [2] Ian Hacking. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press (1983).
- [3] John S. Bell. Against 'Measurement'. *Physics World* 3(8), 33-40 (1990). DOI 10.1088/2058-7058/3/8/26.
- [4] Wojciech H. Zurek. Decoherence, Einselection, and the Quantum Origins of the Classical. *Reviews of Modern Physics* 75(3), 715-775 (2003). DOI 10.1103/RevModPhys.75.715.
- [5] Harry M. Collins. *Changing Order: Replication and Induction in Scientific Practice*. University of Chicago Press (1985).
- [6] JCGM 200:2012. *International Vocabulary of Metrology: Basic and General Concepts and Associated Terms (VIM)*, 3rd edition. BIPM (2012).
- [7] Guido Bacciagaluppi. The Role of Decoherence in Quantum Mechanics. *The Stanford Encyclopedia of Philosophy*, first published 3 November 2003, substantive revision 23 January 2025. <https://plato.stanford.edu/entries/qm-decoherence/>

¿Qué es una señal?

IDE-024 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Seguroola

Versión: v0.9 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: De una tensión por encima del suelo térmico a un pulso que cruzó la galaxia

Límite: La señal como variación físicamente distinguible que porta información sobre un fondo: qué la vuelve detectable, qué la limita y qué no puede señalizarse en absoluto. La metrología del resultado de medida corresponde a IDE-014 y la epistemología del observar a IDE-023

Trabajos relacionados: IDE-023, IDE-019, IDE-018

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Una señal no es lo mismo que un sonido, una luz o una tensión. Es una variación que pudo haber sido otra, seleccionada de un conjunto de posibilidades que el receptor ya trata como posibles, y transportada por un canal que añade ruido propio.[1] La formulación de Shannon lo vuelve preciso y, en el mismo artículo, aparta la cuestión semántica como irrelevante para el problema de ingeniería, un límite que los relatos divulgativos borran con regularidad. La física añade dos restricciones que no son cuestión de técnica: un suelo de ruido fijado por la temperatura y no por la calidad del montaje,[2,3] y una estructura causal que prohíbe de plano ciertas señales, incluidas las que mucha gente espera que permita el entrelazamiento.[4,5] Este artículo pregunta qué debe ocurrir en el mundo para que algo cuente como señal.

En 60 segundos

Una señal es una variación físicamente distinguible tomada de un conjunto de alternativas, y porta información solo en la medida en que estrecha lo que el receptor no sabía ya.[1] Un tono constante, por fuerte que sea, no dice nada después del primer instante. Todo canal añade ruido, y parte de ese ruido no es un defecto del equipo: la agitación térmica de la carga en un conductor produce una fluctuación cuya potencia depende solo de la temperatura y del ancho de banda, medida por Johnson y deducida de la termodinámica por Nyquist el mismo año.[2,3] Por debajo de ese suelo ninguna ingeniería ayuda, aunque una integración más larga y un filtro adaptado a la forma de onda esperada todavía permiten extraer una señal débil.[7] La relatividad prohíbe la influencia controlable fuera del cono de luz, y la teoría cuántica resulta respetarlo incluso para sistemas entrelazados: la estadística de un lado no depende de qué ajuste se elija en el otro, de modo que correlaciones que ningún modelo local reproduce siguen sin transmitir nada.[4,5] Por último, detectar una señal y saber qué la produjo son problemas distintos, y el segundo no lo resuelve el primero.[6]

Por qué importa esta pregunta

Casi toda disputa sobre afirmaciones extraordinarias gira en torno a la diferencia entre una señal y una fluctuación. ¿Aquello fue una detección o una coincidencia? ¿Pudo el efecto haberlo producido el propio aparato? ¿Sería reconocible como tal un mensaje procedente de otro sitio? Estas preguntas tienen

respuestas técnicas, y las respuestas no son intuitivas: el umbral que separa señal de ruido es una elección con consecuencias por ambos lados, el suelo de ruido lo fija la física y no el presupuesto, y algunas transmisiones que suenan a meramente difíciles están prohibidas de plano. Saber cuál es cuál es lo que mantiene a raya tanto el entusiasmo como el desdén.

1. Una diferencia que pudo haber sido otra

Pregunte qué es una señal y la respuesta natural señala un medio: un sonido, un destello, una tensión en un cable. Esa respuesta pierde de vista lo que hace el trabajo. Un tono sostenido a volumen constante es un sonido y no porta nada después del primer momento, porque la información está en el apartamiento respecto de lo esperado y no en la magnitud física misma.

Shannon lo volvió exacto y con ello fundó el campo. La información se mide sobre un conjunto de alternativas con probabilidades asociadas: la cantidad que transmite un resultado concreto depende de cuán improbable era y de cuántas alternativas excluyó.[1] Una fuente que solo puede hacer una cosa no tiene nada que decir, y una fuente cuya salida es completamente impredecible para el receptor transmite el máximo por símbolo.

Shannon especifica un destino y un conjunto de mensajes posibles y exige que el sistema de comunicación funcione para cada selección posible, no solo para el mensaje realmente enviado. Inferencia: en la lectura de ingeniería adoptada aquí, que una fluctuación funcione como mensaje o como ruido depende del sistema receptor y de sus alternativas.

En qué punto estamos

Una señal es una variación tomada de un conjunto de alternativas que el receptor ya trata como posibles, y porta información en proporción a lo que excluye y no en proporción a su energía. Eso hace de la señal una noción relacional y no una propiedad del medio. Lo que no es relacional es cuánta de esa variación aporta el mundo mismo.

2. El suelo es físico y el umbral es una elección

El ruido suele imaginarse como un fallo que la ingeniería debe eliminar. Parte de él lo es: interferencias, captación parásita, un instrumento mal conectado a tierra. Pero una resistencia a temperatura ambiente, sin conectar a nada, produce una tensión fluctuante entre sus bornes, y esa fluctuación no puede eliminarse sin enfriarla.

Johnson midió el efecto en 1928 y mostró que depende de la temperatura y de la resistencia y el ancho de banda del circuito, y no del material ni de los detalles de construcción.[3] Nyquist dedujo el mismo resultado de la termodinámica en un artículo acompañante, y por eso el efecto no es una rareza del cableado de un laboratorio sino una consecuencia del equilibrio entre un conductor y su entorno.[2] Todo amplificador construido desde entonces hereda ese suelo.

La detección por debajo del nivel evidente sigue siendo posible, porque el suelo limita la potencia por unidad de ancho de banda y no lo que puede recuperarse de un registro suficientemente largo. La combinación coherente de observaciones alineadas de una forma de onda conocida y temporizada mejora la relación señal ruido cuando el ruido añadido tiene media cero y no está correlacionado entre observaciones.[7] Un filtro adaptado a una señal conocida maximiza la relación señal-ruido en el instante de muestreo.[7] Un resultado aparte, el de Shannon, da a un canal ruidoso una capacidad: por debajo de ella, la codificación puede hacer la tasa de error tan pequeña como se quiera.[1] Lo que ningún método aporta es certeza. Un umbral bajo para capturar

sucesos débiles acepta falsas alarmas, y uno alto para excluirlas pierde sucesos reales, lo cual es un intercambio y no un problema por resolver.

En qué punto estamos

Parte del ruido de cualquier medida la fijan la temperatura y el ancho de banda, no la calidad del montaje, cosa que Johnson midió y Nyquist dedujo el mismo año. La combinación coherente de observaciones alineadas de una forma de onda conocida y temporizada mejora la relación señal ruido cuando el ruido añadido tiene media cero y no está correlacionado entre observaciones. La física además descarta ciertas transmisiones por completo, sea cual sea el equipo.

3. Qué no puede señalizarse

La restricción que impone la relatividad suele enunciarse como un límite de velocidad, lo que la deja corta. La estructura del espaciotiempo divide los sucesos entre los que pueden influir en uno dado y los que no, y la prohibición recae sobre la influencia controlable: si algún procedimiento permitiera a un experimentador de aquí marcar una diferencia en lo que registra un experimentador de allá fuera del cono de luz, el orden causal de los sucesos dependería del sistema de referencia del observador.[5]

El entrelazamiento parece un modo de sortearlo y no lo es. Las medidas sobre sistemas entrelazados separados están correlacionadas con más fuerza de la que permite cualquier modelo que asigne valores locales preexistentes, y la tentación de leerlo como influencia es comprensible. Lo que da el formalismo, sin embargo, es que las probabilidades de los resultados de un lado no cambian con la elección de ajuste del otro, de modo que la correlación solo se vuelve visible cuando los dos registros se reúnen, cosa que exige un canal ordinario.[5]

Esto es un teorema y no un accidente de los experimentos concretos. Ghirardi, Rimini y Weber mostraron en general que ninguna operación local sobre un lado, es decir ninguna medida ni transformación confinada a ese lado, cambia la estadística accesible en el otro, sea cual sea el estado compartido; de modo que el formalismo cuántico de la medida no puede usarse para transmitir una señal controlable más rápido que la luz.[4] Merece enunciarse con cuidado porque es el punto en el que se equivoca una amplia literatura divulgativa: las correlaciones son reales y no sirven para enviar nada.

En qué punto estamos

La relatividad prohíbe la influencia controlable fuera del cono de luz, y la teoría cuántica cumple: las probabilidades de resultado de un lado de un par entrelazado no dependen del ajuste distante, así que las correlaciones solo se vuelven visibles al comparar registros por un canal ordinario. Eso es un teorema y no una limitación de la técnica actual. Detectar algo real, sin embargo, sigue sin ser lo mismo que saber qué lo envió.

4. Una señal regular no es un mensaje

En 1967 un rastreo de Cambridge registró una fuente de radio que emitía pulsos separados unos 1,34 segundos, repitiéndose con una regularidad distinta de todo lo entonces conocido en un objeto astronómico.[6] La posibilidad de que la fuente fuera artificial se tomó lo bastante en serio como para sopesarla en el artículo del descubrimiento en vez de descartarla.

El razonamiento que lo zanjó no trataba de la regularidad sino de todo lo demás, y sus hebras hacían trabajos distintos. La ausencia de paralaje medible situaba la fuente mucho más allá del

sistema solar, mientras que la deriva del tiempo de llegada con la frecuencia era compatible con la dispersión en el plasma interestelar y, bajo el modelo empleado entonces, implicaba un recorrido del orden de 65 parsecs. Los tiempos de llegada tampoco mostraban movimiento orbital adicional comparable con el de la Tierra, lo que argumentaba contra un transmisor en un planeta en órbita. [6] La hipótesis del mensaje no sobrevive a ninguna de las tres.

La lección general es que detección e identificación son problemas distintos apoyados en evidencia distinta. Un patrón que pudo haber sido otro es una señal según la definición de ingeniería, sea lo que sea que lo produjo, y estrechar el origen exige encontrar restricciones independientes, que es lo que aportaron el paralaje, la dispersión y los tiempos de llegada. La regularidad por sí sola es evidencia débil sobre el origen, porque el universo contiene relojes excelentes.

En qué punto estamos

El primer púlsar era una señal según todo criterio de ingeniería y su origen seguía abierto, zanjado no por su regularidad sino por una cota de paralaje, por la dispersión en el plasma interestelar y por la ausencia de movimiento orbital más allá del de la Tierra. Detección e identificación piden evidencia distinta. Lo que ninguna de las dos aporta es qué significa la señal.

5. Dónde acaba la señal y empieza el significado

Shannon abrió su artículo apartando el significado: los aspectos semánticos de la comunicación, escribió, son irrelevantes para el problema de ingeniería, que consiste en reproducir en un punto, exacta o aproximadamente, un mensaje seleccionado en otro.[1] La restricción era deliberada y es la razón por la que la teoría funciona. A los bits no les importa de qué tratan.

La restricción también se olvida con regularidad. Las afirmaciones de que un sistema transmite información, o de que un genoma o una neurona portan un mensaje, hacen a menudo la afirmación de ingeniería, que es medible, mientras suenan a la semántica, que es una afirmación distinta y necesita su propio argumento. Mantenerlas separadas no cuesta nada y evita muchísima confusión.

Lo que queda tras la restricción es una imagen física limpia. Algo varía de un modo que pudo haber sido otro; la variación cruza un canal que añade ruido con un suelo fijado por la temperatura; un receptor que sostiene un conjunto de hipótesis lo registra y decide, bajo un umbral que intercambia errores de dos clases; y la pregunta por qué lo produjo, y por de qué trata, la responde evidencia enteramente distinta.

En qué punto estamos

La teoría de Shannon funciona porque excluye el significado, y confundir la afirmación de ingeniería con la semántica es el error más común cometido en su nombre. Una señal es una variación que pudo haber sido otra, que sobrevive al ruido y se registra contra una hipótesis. De qué es evidencia es la pregunta que trae el observador, y responderla es de lo que ha tratado el resto de esta serie.

Lo que sabemos y cómo lo sabemos

- Una señal es una variación tomada de un conjunto de alternativas, y la información que porta depende de lo que excluye y no de su energía.[1]
- Parte del ruido de cualquier medida eléctrica es térmico y depende de la temperatura, la resistencia y el ancho de banda, no de la construcción; se midió y se dedujo de forma independiente en 1928.[2,3]

- Ningún procedimiento que use sistemas entrelazados transmite una señal controlable, porque las probabilidades de resultado de un lado son independientes del ajuste elegido en el otro.[4,5]
- Detectar una señal e identificar qué la produjo son problemas distintos apoyados en evidencia distinta, como mostró el primer púlsar.[6]

Lo que todavía no sabemos

- Si la restricción de no señalización sobreviviría en una teoría que sustituyera a la mecánica cuántica, dado que las demostraciones disponibles la establecen dentro del formalismo actual y no como principio independiente.[4,5]
- Cómo enunciar, sin colar los propósitos de un receptor, cuándo una variación física cuenta como portadora de información acerca de algo en vez de estar meramente correlacionada con ello, que es la cuestión que Shannon apartó.[1]
- Qué umbral es apropiado para anunciar una detección en campos donde la probabilidad previa del efecto es genuinamente desconocida, cosa que la matemática de la decisión bajo ruido no aporta por sí sola.[1]

Qué podría hacernos cambiar de opinión

- Un protocolo reproducible con sistemas entrelazados que cambiara la estadística de resultados de un lado según un ajuste elegido en el otro contradiría el resultado de no señalización y, con él, la compatibilidad de la teoría cuántica con el orden causal relativista.[4,5]
- Una medida de la potencia de ruido eléctrico en equilibrio térmico que se apartara sistemáticamente de su dependencia de la temperatura, la resistencia y el ancho de banda contradiría la deducción termodinámica del suelo de ruido.[2,3]
- Una fuente de pulsos astronómicos cuya dispersión y cuyos tiempos de llegada coincidieran con un transmisor en un cuerpo en órbita y no con un objeto natural reabriría la cuestión de identificación que el caso de los púlsares se considera haber zanjado.[6]

Si solo recuerdas una cosa

Una señal es una diferencia que pudo haber sido otra, que sobrevive a un canal ruidoso y la reconoce un receptor que ya sabía qué buscar.

Artículos relacionados

- IDE-023 — ¿Qué es la observación?
- IDE-019 — ¿Qué es la luz? Onda, cuanto y observador
- IDE-018 — Conexión: qué afirma una arista

Referencias

[1] Claude E. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423 and 623-656 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x; 10.1002/j.1538-7305.1948.tb00917.x.

[2] Harry Nyquist. Thermal Agitation of Electric Charge in Conductors. Physical Review 32, 110-113 (1928). DOI 10.1103/PhysRev.32.110.

[3] John B. Johnson. Thermal Agitation of Electricity in Conductors. Physical Review 32, 97-109 (1928). DOI 10.1103/PhysRev.32.97.

[4] GianCarlo Ghirardi, Alberto Rimini and Tullio Weber. A General Argument against Superluminal Transmission through the Quantum Mechanical Measurement Process. Lettere al Nuovo Cimento 27, 293-298 (1980).

- [5] Asher Peres and Daniel R. Terno. Quantum Information and Relativity Theory. *Reviews of Modern Physics* 76(1), 93-123 (2004). DOI 10.1103/RevModPhys.76.93.
- [6] Antony Hewish, S. Jocelyn Bell, John D. H. Pilkington, Paul F. Scott and Robin A. Collins. Observation of a Rapidly Pulsating Radio Source. *Nature* 217, 709-713 (1968). DOI 10.1038/217709a0.
- [7] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley (1968), chapter 4.