

La unidad que falta: la falacia mereológica y el argumento del sustrato

Autor: Steven Vallejo Ortiz **Curso:** Filosofía de las Neurociencias (2026-1) **Profesor:** Santiago Arango-Muñoz **Institución:** Instituto de Filosofía, Universidad de Antioquia

Introducción

Bennett y Hacker (2022) señalaron un error preciso: atribuir a una parte lo que se predica del todo (la *falacia mereológica*). El cerebro no ve; ve la persona. Este ensayo aplica esa lección a un problema: ¿podría una máquina de silicio ser consciente? Frente al funcionalismo, que responde que sí, se ha consolidado una réplica —la conciencia pertenece a los seres vivos; lo vivo se autoproduce; el silicio no, luego no siente—. La llamaré *el argumento del sustrato*.

El debate vive de confundir dos tesis vecinas. **No niego que la vida sea condición necesaria de la conciencia:** es plausible. Niego que la auto-producción sea el **criterio que individúa al sujeto**, lo que decide cuál de los sistemas anidados en un cuerpo siente. Una condición no es un criterio, y el argumento necesita ambos: sin el segundo, de «el silicio no se autoproduce» nada se sigue sobre lo que pueda sentir. La autopoiesis es nítida pero señala la unidad equivocada: individúa **células**, no personas.

1. El problema y su reconstrucción

La metáfora del cerebro como computadora acaba dictando la teoría (Daugman, 2001). Su versión filosófica es la realizabilidad múltiple (Putnam, 1967; Fodor, 1974): los estados mentales se definen por rol causal, no por composición, y son realizables en cualquier medio apto. Corolario: el sustrato sería ontológicamente irrelevante.

Contra eso convergen dos tradiciones. Searle (1980) sostiene, con la habitación china, que la sintaxis no basta para la semántica, y su naturalismo biológico (1992), que la conciencia es un fenómeno causado por procesos neurofisiológicos. La segunda, de la biología teórica: Maturana y Varela definen lo vivo por la **autopoiesis**, un sistema que produce y regenera componentes y frontera (Varela, Maturana y Uribe, 1974; Maturana y Varela, 1980). Thompson (2007) extrae la continuidad

vida-mente; Di Paolo (2005) la corrige desde dentro: hace falta **adaptividad**, la regulación activa de la propia precariedad.

Del lado neurocientífico, la interocepción muestra cuánta experiencia se organiza alrededor de la regulación corporal (Chen et al., 2021; Greenwood y Garfinkel, 2025). Seth (2025) lleva eso a su conclusión filosófica: la conciencia es «una propiedad únicamente de los sistemas vivos (aunque no necesariamente de todos)»; sus argumentos, dice, «no refutan» el funcionalismo, «pero lo vuelven menos plausible», y su eje es vivo/no-vivo, no carbono/silicio.

El argumento canónico: la conciencia requiere un sujeto, un sistema uno cuyos estados tengan valencia; un sistema tiene valencia sólo si se produce a sí mismo y puede morir; ningún sistema de silicio se produce a sí mismo; luego ninguno es sujeto de conciencia.

Las dos primeras premisas exigen a la autopoiesis que no sólo distinga lo vivo de lo no vivo, sino que **individúe al portador de la experiencia**: que el sujeto esté donde está la auto-producción. Ahí se rompe.

2. La unidad equivocada

La unidad canónica autopoietica es **la célula**: paradigma de Varela, Maturana y Uribe (1974). Es objetiva: hay un hecho sobre si una red metabólica produce su membrana.

El problema es de aritmética. En mi cuerpo hay del orden de 10^{13} unidades autopoieticas, y **ninguna soy yo**. Si el sujeto está donde está la auto-producción, hay diez billones de sujetos y ninguno escribe este ensayo. El argumento que debía excluir al silicio termina multiplicando sujetos dentro del carbono.

De ahí tres consecuencias. Primera: **el cerebro no es una unidad autopoietica**. No produce sus componentes; los producen sus células. Es un sistema de *segundo orden*: una unidad de unidades autopoieticas que no es autopoietica.

Segunda: **el hígado es tan autopoietico como el cerebro**. Si la auto-producción individuara al sujeto, nada explicaría por qué es el cerebro y no el hígado. Un criterio que no distingue entre órganos no vale.

Tercera: para escapar hay que añadir —integración causal, clausura operacional— algo que seleccione la escala. Pero **eso añadido hace todo el trabajo**: es una propiedad de la organización, no del material, lo que el funcionalista concede (Mylopoulos, 2022). La auto-producción queda

como un epiciclo.

La neurociencia del curso enseña dónde sí está el correlato. Quian Quiroga, Fried y Koch (2013) hallaron neuronas que responden a una persona, pero su lectura es distribuida: cada concepto tendría «un conjunto de neuronas correspondientes asignadas a él». Cada una se autoproduce igual que las demás; lo que varía —y explica— es con quién está conectada. La auto-producción es constante en todo el cuerpo; el correlato no.

La falacia mereológica reaparece invertida: el argumento elige un criterio que sólo vale para las partes —las células— y pretende decidir sobre el todo. El grillo de Webb (1996) delega la fonotaxis a un tubo de desfase, materia no viva y cognitivamente constitutiva.

3. La clínica ya separó las variables

No hace falta esperar una máquina consciente: la neurología ya disoció en pacientes las dos variables que importan. Laureys (2007) describe el estado vegetativo como la disociación de los dos componentes de la conciencia: «la vigilia permanece intacta, pero la consciencia —que abarca todos los pensamientos y sentimientos— queda abolida». Estos pacientes conservan ciclos de sueño-vigilia, respiran sin asistencia y hacen movimientos espontáneos. La auto-producción está **completa**: el organismo se automantiene, está vivo. Y el sujeto no está.

Laureys añade algo más incómodo: algunos «voluntarios sanos y plenamente conscientes tienen valores de metabolismo cerebral global comparables a los de algunos pacientes en estado vegetativo». Concluye: «medir los niveles globales de consumo de energía en el cerebro no puede indicar la presencia de consciencia». Lo que sí covaría con el sujeto es la «conectividad funcional dentro de esta red frontoparietal y con centros más profundos del cerebro, en particular el tálamo». En sus pacientes el estímulo llegaba a la corteza somatosensorial primaria y allí moría: «estaba aislada y desconectada del resto del cerebro».

El tejido está vivo, se autoproduce y responde. Lo que falta es **integración**. Laureys documenta conciencia encubierta detectada por neuroimagen y diagnóstico errado en más de un tercio de los casos; la cautela obliga a hablar de la categoría clínica, y la inferencia es mía. Pero basta: la clínica mantiene constante la auto-producción y mueve al sujeto. El factor diferencial **no es la auto-producción** —es organizacional—.

4. La tradición ya lo sabía

No es un descubrimiento mío: es la razón por la que Varela abandonó el marco estrecho. En *Principles of Biological Autonomy* (1979) generaliza la autopoiesis a la **autonomía**, articulada por la **clausura operacional**, porque la autopoiesis estricta es celular y no llega al organismo ni al sistema nervioso; la autonomía sí, pues no exige producción material, sino una red de procesos recíprocos.

La clausura operacional es **organizacional**: se define por cómo se encadenan los procesos, no por de qué están hechos. Una propiedad de la forma de la red es realizable en más de un medio. O la tradición se queda en la autopoiesis estricta —que no individúa al sujeto—, o asciende a la autonomía —que lo individúa, sin excluir sustrato—. No hay tercera posición. **El camino que el enactivismo recorrió para individuar al sujeto es el mismo que le quita el argumento contra el silicio.**

Beer (2004), analizando un *glider* del juego de la vida, observa que su identidad «depende crucialmente de cómo elegimos definir» su organización. El criterio es nítido en el caso paradigmático e indeterminado donde el argumento lo necesita.

5. Lo que construí, y lo que obtuve en su lugar

Construí un experimento por cada lectura del seminario: la jerarquía visual de Zeki (1992), las células de concepto de Quian Quiroga, la neuroquímica de LeDoux (1994), la plasticidad de Hinton (1992), el grillo de Webb. Ninguno encontró la diferencia de sustrato que yo buscaba. Los que sobrevivieron a mi auditoría me devolvieron la constante que había escrito antes de correrlos; otros ni siquiera llegaron a eso.

Mi única medición genuina fue un benchmark de diecinueve configuraciones, y es la que más me enseñó, porque empecé buscando en ella la prueba de que el silicio es energéticamente inviable. La brecha de eficiencia entre carbono y silicio recorre, en mis propios datos, de $2,9 \times 10^3$ a $4,1 \times 10^5$ —se mueve ciento cuarenta veces— y no monotónicamente: el mínimo no está en ningún extremo, sino en medio, y repartir el mismo cómputo entre dos tarjetas la **empeora**. Un número que cambia dos órdenes de magnitud según cómo distribuyo el trabajo no mide una propiedad del silicio: mide mi ingeniería. Laureys ya lo había dicho desde la clínica, y con pacientes: el consumo de energía no

indica la presencia de consciencia. Mi intuición no era falsa; estaba mal dirigida. Medía con cuidado la magnitud equivocada.

Nada de esto prueba mi tesis, y conviene decirlo antes de que lo diga otro: un modelo mal construido siempre devuelve su estipulación, y eso es un hecho sobre mí, no sobre el mundo. Un refutador de lo que sostengo sería una medida de la firma autopoietica **invariante ante el corte**: que diera igual contando o no la fuente de alimentación. Eso intenté construir —un coeficiente κ de acoplamiento entre cómputo y automantenimiento—, y la auditoría mostró que fallaba por aritmética: el parámetro lesionado no aparece siquiera en la ecuación. El argumento lo llevan las secciones anteriores, no mi código. Pero sé qué obtuve en su lugar, cada vez: **el corte**. Con la máquina intacta y la misma perturbación, contar la fuente de alimentación como parte del sistema en vez de como un exterior inagotable lleva κ de 0,00 a 0,70. No cambió el sustrato: cambió dónde tracé la línea.

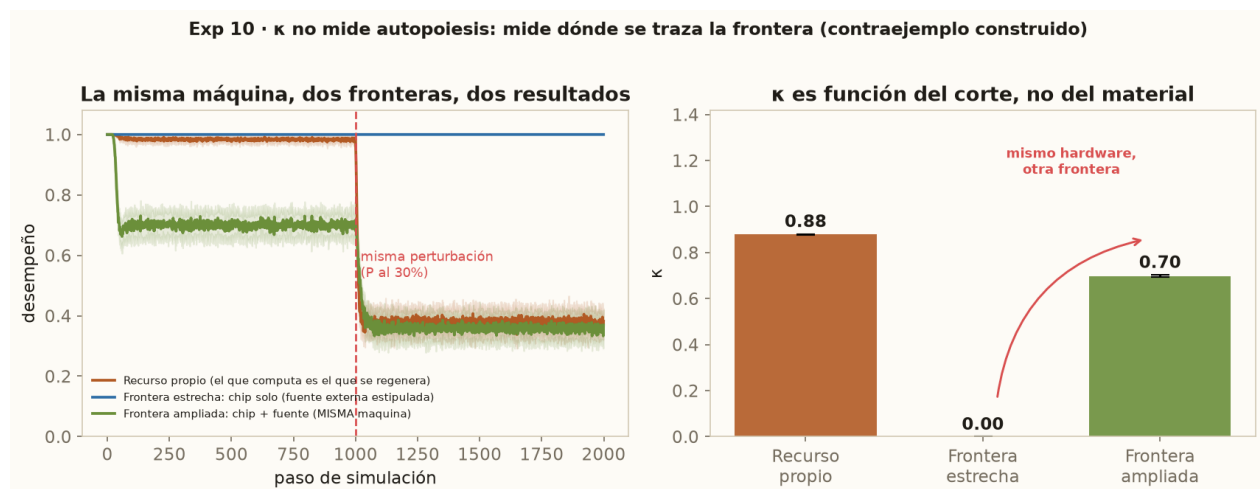


Figura 1: *Experimento 10 (reformado)*. La misma máquina, bajo la misma perturbación, con dos fronteras distintas: contando sólo el chip, $\kappa = 0,00$ (inmune); contando también su fuente de alimentación, $\kappa = 0,70$ (24 semillas). El sustrato no cambió: cambió el corte.

6. La objeción más fuerte, y los límites

La réplica seria es: «de acuerdo, la auto-producción no individúa al sujeto; pero sigue siendo **necesaria**. El silicio no se autoproduce, luego le falta una condición sin la cual no hay consciencia».

Concedo la premisa y niego la inferencia. ¿De qué es necesaria la autopoiesis? Si del sujeto, el sujeto tendría que ser una unidad autopoietica, y ninguno lo es: somos sistemas de segundo orden.

Si de los componentes, es una tesis sobre el material, y hace falta un argumento que diga por qué ese material y no otro; argumento que la autopoiesis no puede dar: habla de células, no de sujetos. Una condición de fondo no es un criterio de exclusión: que todo lo consciente conocido sea vivo no implica que sólo lo vivo pueda serlo.

Reconozco de quién es lo que sobrevive. La distinción entre identidad propia y prestada —la del artefacto, cuya identidad «le es acordada» por otro— es de Jonas (1968). Que el problema de la individualidad sea «el de justificar cuál separación elegimos entre el gran conjunto de distinciones posibles y arbitrarias» está en Barandiaran, Di Paolo y Rohde (2009). Y Villalobos y Dewhurst (2018) demostraron que una máquina de Turing físicamente implementada satisface la clausura operacional del enactivismo clásico, que por tanto «no necesita descartar las caracterizaciones computacionales del sistema nervioso». Míos son el primer cuerno —que la autopoiesis estricta individúa células— y la forma de tenaza.

No he mostrado que una máquina sienta, ni que la vida sea irrelevante para la mente. Tampoco he explicado los qualia; mi argumento es neutral ante el problema difícil.

Conclusión

El argumento del sustrato parecía tener dos cosas: un criterio material —la auto-producción— y un sujeto al cual aplicarlo. Tiene sólo una a la vez: en sentido estricto señala células —diez billones en un cuerpo, ninguna de ellas yo—; en el amplio individúa organismos, pero se ha vuelto una propiedad de la organización que no excluye sustratos: es la tesis del adversario.

Bennett y Hacker enseñaron que atribuir a la parte lo del todo fabrica pseudoproblemas. El argumento del sustrato es un caso: discute de qué está hecho el sujeto antes de decir cuál es. Queda lo que perseguía mi laboratorio, y no donde lo buscaba: la diferencia real no es carbono contra silicio, sino automantenimiento **delegado** contra **constitutivo** —un sistema cuyas condiciones de persistencia son internas frente a uno que las tiene fuera, en técnicos y redes eléctricas cuya normatividad es ajena—. Ahí sí importa la energía: no por cuánta se gasta, sino por quién responde de ella. Pero esa diferencia es relativa al corte, y decidirlo es la tarea que el debate ha estado saltándose. Mientras la respuesta la ponga quien pregunta, «¿puede el silicio ser consciente?» no es todavía una pregunta: es un recorte a la espera de un argumento.

Bibliografía

Fuentes primarias

- Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). *Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action*. *Adaptive Behavior*, 17(5), 367-386.
- Beer, R. D. (2004). *Autopoiesis and Cognition in the Game of Life*. *Artificial Life*, 10(3), 309-326.
- Di Paolo, E. A. (2005). *Autopoiesis, Adaptivity, Teleology, Agency*. *Phenomenology and the Cognitive Sciences*, 4(4), 429-452.
- Fodor, J. A. (1974). *Special Sciences (Or: The Disunity of Science as a Working Hypothesis)*. *Synthese*, 28(2), 97-115.
- Jonas, H. (1968). *Biological Foundations of Individuality*. *International Philosophical Quarterly*, 8(2), 231-251.
- Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.
- Putnam, H. (1967). *Psychological Predicates*. En W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion*. University of Pittsburgh Press.
- Searle, J. R. (1980). *Minds, Brains, and Programs*. *Behavioral and Brain Sciences*, 3(3), 417-457.
- Searle, J. R. (1992). *The Rediscovery of the Mind*. MIT Press.
- Seth, A. K. (2025). *Conscious artificial intelligence and biological naturalism*. *Behavioral and Brain Sciences*, publicación anticipada en línea, 1-42. <https://doi.org/10.1017/S0140525X25000032>
- Varela, F. J. (1979). *Principles of Biological Autonomy*. North-Holland. [Ed. anotada: E. A. Di Paolo & E. Thompson (Eds.), MIT Press, 2024.]
- Varela, F. J., Maturana, H. R., & Uribe, R. (1974). *Autopoiesis: The Organization of Living Systems, Its Characterization and a Model*. *BioSystems*, 5(4), 187-196.
- Villalobos, M., & Dewhurst, J. (2018). *Enactive autonomy in computational systems*. *Synthese*, 195(5), 1891-1908. <https://doi.org/10.1007/s11229-017-1386-z>
- Webb, B. (1996). *A Cricket Robot*. *Scientific American*, 275(6), 94-99.

Fuentes secundarias

- Bennett, M. R., & Hacker, P. M. S. (2022). *The Mereological Fallacy in Neuroscience*. En *Philosophical Foundations of Neuroscience* (2ª ed.). Wiley-Blackwell.
- Chen, W. G., y cols. (2021). *The emerging science of interoception: Sensing, integrating, interpreting, and regulating signals within the self*. *Trends in Neurosciences*, 44(1), 3-16. <https://doi.org/10.1016/j.tins.2020.10.007>
- Daugman, J. (2001). *Brain Metaphor and Brain Theory*. En W. Bechtel, P. Mandik & J. Mundale (Eds.), *Philosophy and the Neurosciences: A Reader*. Blackwell.
- Greenwood, B. M., & Garfinkel, S. N. (2025). *Interoceptive mechanisms and emotional processing*. *Annual Review of Psychology*, 76(1), 59-86. <https://doi.org/10.1146/annurev-psych-020924-125202>
- Hinton, G. E. (1992). *How Neural Networks Learn from Experience*. *Scientific American*, 267(3), 144-151.
- Laureys, S. (2007). *Eyes Open, Brain Shut*. *Scientific American*, 296(5), 84-89.
- LeDoux, J. E. (1994). *Emotion, Memory and the Brain*. *Scientific American*, 270(6), 50-57.
- Mylopoulos, M. (2022). *Neurobiological theories of consciousness*. En B. D. Young & C. D. Jennings (Eds.), *Mind, Cognition, and Neuroscience: A Philosophical Introduction* (pp. 280-293). Routledge.
- Quian Quiroga, R., Fried, I., & Koch, C. (2013). *Brain Cells for Grandmother*. *Scientific American*, 308(2), 30-35.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Zeki, S. (1992). *The Visual Image in Mind and Brain*. *Scientific American*, 267(3), 68-76.