

Índice

El argumento del sustrato y la unidad que le falta	1
Resumen	1
0. Introducción y hoja de ruta	2
Parte I. El argumento del sustrato en su mejor forma	3
Parte II. La unidad equivocada	6
Parte III. La clínica ya separó las variables	11
Parte IV. La tradición ya lo sabía	13
Parte V. El laboratorio: qué mide y qué no	16
Parte VI. Objeciones y límites	22
Conclusión	28
Glosario	29
Bibliografía	32

El argumento del sustrato y la unidad que le falta

La falacia mereológica, la autopoiesis y el problema de la individuación en el debate sobre la conciencia artificial

Autor: Steven Vallejo Ortiz **Curso:** Filosofía de las Neurociencias (2026-1), Instituto de Filosofía, Universidad de Antioquia **Profesor:** Santiago Arango-Muñoz **Documento:** versión extendida del ensayo, con la auditoría del laboratorio

Resumen

Examino un solo problema: si la autopoiesis puede decidir qué sistemas pueden ser conscientes. Reconstruyo primero, con la máxima caridad, el argumento que llamo *del sustrato*: la conciencia pertenece a los seres vivos; lo vivo se define por producirse a sí mismo; el silicio digital no se produce a sí mismo; luego no siente. Sostengo que ese argumento le exige a la autopoiesis algo que la autopoiesis no entrega: **individuación al portador de la experiencia**. Y no falla por vaguedad —la autopoiesis es objetiva, y fijada una granularidad es decidible—, sino por **extensión**: individúa células. En un cuerpo humano hay del orden de 10^{13} unidades autopoieticas y ninguna es el sujeto; el cerebro no es una unidad autopoietica, sino un sistema de segundo orden; y el hígado es tan

autopoiético como el cerebro. Para seleccionar al sujeto entre escalas anidadas hay que añadir integración causal, que es sustrato-neutral. La propia tradición lo advirtió: Varela (1979) generalizó de la autopoiesis a la autonomía y la clausura operacional precisamente porque la autopoiesis estricta es celular. Pero la clausura operacional es organizacional, y lo organizacional es realizable en más de un medio. El camino que el enactivismo recorrió para individuar al sujeto es el mismo que le retira el argumento anti-funcionalista. La neurología clínica corrobora el diagnóstico sin proponérselo: en el estado vegetativo la auto-producción del organismo está completa y el sujeto, en la categoría clínica así definida, no está; lo que covaría con su presencia es la conectividad frontoparietal y córtico-talámica, no el metabolismo, que Laureys (2007) descalifica explícitamente como indicador. Documento además la auditoría de un laboratorio propio cuyo resultado fue negativo. No se sigue que el silicio sienta; se sigue que la autopoiesis no es el instrumento que puede decidirlo.

0. Introducción y hoja de ruta

Bennett y Hacker (2022) le enseñaron a la filosofía de las neurociencias a desconfiar de un error preciso: atribuir a una parte lo que sólo se predica del todo. El cerebro no ve, no decide ni cree; ve, decide y cree la persona. Llamen a esto la *falacia mereológica*, y su lección de método —la que me interesa aquí— es que ciertas disputas no se resuelven acumulando datos, porque el defecto no está en la evidencia sino en el sujeto de la oración. Antes de preguntar qué propiedad tiene X, hay que haber fijado cuál es X.

Este documento aplica esa lección a un problema único: **¿puede la autopoiesis decidir qué sistemas pueden ser conscientes?** Mi respuesta es que no, y quiero ser exacto de entrada sobre el estatus de esa negación, porque casi todo el malentendido posible se juega ahí. No afirmo que una máquina de silicio sea o pueda ser consciente. No afirmo que la vida sea irrelevante para la mente. No afirmo que la autopoiesis sea una noción confusa, ni relativa al observador, ni científicamente ociosa. Afirmo algo más estrecho y, creo, más firme: que el argumento que va de «el silicio no se autoproduce» a «el silicio no siente» necesita, para ser válido, que la auto-producción material **individúe al sujeto** de la conciencia, y que no hace eso. Individúa células.

Llamaré a ese argumento **el argumento del sustrato**. Es una reconstrucción mía —su forma

canónica la enuncio en la Parte I—; no se la atribuyo verbatim a ningún autor, aunque sus piezas están en autores reales y las cito.

La hoja de ruta: la **Parte I** reconstruye el argumento en su mejor forma, antes de tocarlo, porque un argumento derrotado en su versión débil no se ha derrotado. La **Parte II** presenta el cargo central: la autopoiesis individúa la unidad equivocada, y el cargo es de extensión, no de vaguedad. La **Parte III** muestra que la clínica ya separó, en pacientes reales, las dos variables que el argumento funde: en el estado vegetativo la auto-producción está completa y el sujeto no está. La **Parte IV** muestra que la tradición ya conocía el problema de escala y se movió por él —Varela (1979) generaliza de la autopoiesis a la autonomía— sin notar que ese movimiento le retira el argumento anti-funcionalista que quería fundar. La **Parte V** audita un laboratorio propio: construí una medida de la «firma operacional» de la autopoiesis y, auditada, medía otra cosa. La **Parte VI** enfrenta las objeciones serias —empezando por la más fuerte, que la vida siga siendo condición necesaria aunque no individúe— y fija dónde se detiene el argumento. Cierran una conclusión, un glosario que distingue sin equiparar las nociones que el debate suele fundir, y la bibliografía.

Parte I. El argumento del sustrato en su mejor forma

I.1. El corolario que provoca la réplica

La metáfora del cerebro como computadora tiene una historia y un precio, y Daugman (2001) documentó ambos: cada época piensa el cerebro con su tecnología dominante —la bomba hidráulica, el telégrafo, la centralita, el ordenador digital—, y la metáfora, que empieza siendo una ayuda heurística, acaba dictando la teoría. Su versión filosóficamente disciplinada es la **realizabilidad múltiple** (Putnam, 1967; Fodor, 1974): los estados mentales son estados funcionales, individuados por su rol causal —por lo que hacen, qué los causa y qué causan— y no por aquello de lo que están hechos. De ahí el corolario que provoca toda la discusión: si el tipo mental se define por el rol, cualquier medio físico capaz de sostener las transiciones adecuadas lo realiza. El sustrato sería ontológicamente irrelevante.

Contra ese corolario convergen dos tradiciones que conviene no confundir.

I.2. La primera vertiente: Searle

Searle (1980) sostiene, con la habitación china, que la manipulación formal de símbolos según reglas no basta para la semántica: ejecutar el programa correcto no es, por sí mismo, comprender. Su naturalismo biológico (1992) desarrolla el flanco positivo: la conciencia es un fenómeno biológico real, causado por procesos neurofisiológicos y realizado en el sistema nervioso, como la digestión es causada por procesos y realizada en un aparato. No es una propiedad del programa, sino del sistema físico que lo ejecuta, y no cualquier sistema físico la tiene.

Nótese qué le falta a Searle para lo que me ocupa: nos dice que la conciencia es causada biológicamente, pero no entrega un criterio positivo, operativo e independiente que permita señalar **qué** sistemas tienen los «poderes causales» adecuados. Ese hueco es el que la segunda vertiente promete llenar.

I.3. La segunda vertiente: la autopoiesis y la continuidad vida-mente

La segunda vertiente viene de la biología teórica. Maturana y Varela definen lo vivo no por una lista de rasgos (metabolismo, reproducción, irritabilidad) sino por una **organización**: un sistema autopoietico es una red de procesos de producción de componentes que, mediante sus interacciones, regenera continuamente la red que los produjo y realiza la frontera que la constituye como unidad discreta (Varela, Maturana y Uribe, 1974; Maturana y Varela, 1980). La célula es el caso paradigmático y el que modelan: la red metabólica produce la membrana, y la membrana encierra y hace posible la red metabólica. El sistema es su propio producto; la circularidad no es un defecto lógico, es la tesis.

De aquí Thompson (2007) extrae la **tesis de continuidad vida-mente**: los principios organizacionales que hacen de un sistema un ser vivo son los mismos que, elaborados, hacen de él un ser cognitivo. Y Di Paolo (2005) corrige a la tradición desde dentro, con una objeción que considero decisiva y que usaré: la autopoiesis *desnuda* no basta. Un sistema meramente autopoietico o se conserva o se desintegra; no tiene grados, y por tanto nada puede *importarle*. Para que haya normatividad intrínseca —para que algo sea mejor o peor **para el sistema mismo**— hace falta **adaptividad**: regular activamente la propia posición respecto de las condiciones de viabilidad, registrar la tendencia y actuar antes del colapso. Es la adaptividad, y no la autopoiesis sola, la que

engendra valencia.

I.4. El respaldo neurocientífico

Del lado empírico la convergencia es real, y el seminario la recorrió. La ciencia de la **interocepción** —el sentido del estado del propio cuerpo: señales viscerales, cardiovasculares, inmunes, metabólicas— muestra hasta qué punto la experiencia se organiza alrededor de la regulación del medio interno (Chen et al., 2021; Greenwood y Garfinkel, 2025). El afecto no aparece como una capa expresiva sobre una cognición neutra, sino como el modo en que el sistema registra su propia condición de viabilidad. Seth (2021) construye sobre esto su tesis del *beast machine*: percibimos el mundo, y a nosotros mismos, con y a través de un cuerpo cuya persistencia hay que asegurar.

Seth (2025) lleva esto a su formulación más cuidada, y aquí quiero ser escrupuloso, porque es el defensor más fuerte de la posición que critico y sería fácil ganarle atribuyéndole lo que no dice. Su tesis es que «la conciencia es una propiedad únicamente de los sistemas vivos (aunque no necesariamente de todos)». Su argumento nuclear es que simular algo no es realizarlo: «nada se moja dentro de una computadora que predice el clima». Pero acota su pretensión con una honestidad que hay que devolverle: «los argumentos presentados hasta aquí no refutan el funcionalismo computacional. Pero lo vuelven menos plausible y menos atractivo». Y rechaza explícitamente el eje que suele atribuírsele: «tampoco suscribo el *chauvinismo* del carbono (la afirmación de que sólo la vida basada en carbono puede ser consciente)». Su eje **no es carbono/silicio; es vivo/no-vivo**. Toda crítica que le impute carbono-centrismo yerra el blanco, y la mía no se lo imputa.

I.5. El argumento en forma canónica

Reunidas las piezas, la posición que me interesa puede enunciarse así:

(P1) La conciencia requiere un **sujeto**: un sistema que sea *uno* —una unidad genuina y no un agregado que nosotros recortamos— y para el cual sus propios estados tengan **valencia**, es decir, importen desde su propio punto de vista.

(P2) Un sistema tiene valencia propia sólo si es **precario** y se **produce a sí mismo**: sólo si su existencia es una tarea que él mismo ejecuta y que puede fracasar. La identidad

que no se produce a sí misma le es **acordada** desde fuera (Jonas, 1968).

(P3) Ningún sistema de silicio digital se produce a sí mismo: sus componentes los fabrican otros, su persistencia la garantizan técnicos y redes eléctricas, y su normatividad —qué cuenta como funcionar bien— es previa y ajena.

(C) Luego ningún sistema de silicio digital es sujeto de conciencia.

Insisto en que ésta es la versión **fuerte** del argumento, no un muñeco de paja. (P2) no es una estipulación caprichosa: llena el hueco que Searle deja abierto con un criterio positivo, material y objetivo. (P3) es, en el estado actual de la técnica, sencillamente verdadera, y no la disputo. Y la conclusión no es una intuición: es la consecuencia de dos premisas que muchos filósofos serios aceptan.

I.6. Qué le exige el argumento a la autopoiesis

Aquí está el punto que organiza todo lo que sigue. (P1) y (P2) **conjuntamente** no le exigen a la autopoiesis distinguir lo vivo de lo no vivo: eso lo hace, y bien. Le exigen algo mucho más fuerte: **indivduar al portador de la experiencia**. (P1) postula un sujeto; (P2) dice que la marca del sujeto es la auto-producción. Juntas afirman, no que los seres vivos tengan una propiedad interesante, sino que **el sujeto está donde está la auto-producción**. La auto-producción no funciona ahí como un rasgo que el sujeto exhibe: funciona como el **principio de individuación** que dice cuál es el sujeto.

Ésa es la exigencia que el argumento no puede pagar. Y quiero señalar dónde se rompe, porque casi todo el debate lo busca en el lugar equivocado: no se rompe en (P3), la premisa sobre el silicio, que concedo. Se rompe en el paso de (P1) a (P2), en el supuesto tácito de que la auto-producción individúa sujetos. Individúa unidades, sí. Pero no las que hacen falta.

Parte II. La unidad equivocada

II.1. La extensión de «sistema autopoiético» es la célula, y es objetiva

Preguntemos, con toda literalidad, por la **extensión** del predicado «es un sistema autopoiético». ¿De qué se predica, en el mundo?

La respuesta de la tradición es unívoca: la unidad canónica es **la célula**. Es el caso que Varela, Maturana y Uribe (1974) modelan explícitamente, el ejemplo con el que la teoría se enseña, el paradigma al que vuelve cada exposición. Y es el único caso en el que las tres condiciones se satisfacen sin discusión: la red produce los componentes, la red produce la frontera, y la frontera es condición de la red.

Y es objetiva. Quiero decirlo con la máxima claridad, porque una versión anterior de mi propio argumento cometía aquí un error grave y quiero dejarlo enterrado. No acuso a la autopoiesis de vaguedad ni de relatividad al observador. Fijada una granularidad —fijado, por ejemplo, que «componente» significa macromolécula—, la pregunta «¿esta bacteria produce los componentes que la producen?» tiene una **respuesta física objetiva**, independiente de nosotros, y la biología la responde. Que las definiciones tengan un momento convencional en su fijación no vuelve subjetivas a las propiedades que definen; si así fuera, ninguna ciencia tendría objetos. Argüir de la convencionalidad de una definición a la irrealdad de la propiedad es el error de decir «la vida es ambigua en los virus, luego la vida no existe». No lo cometo.

Mi cargo es de otro tipo: es un cargo de **extensión**. El predicado es nítido; lo que ocurre es que su extensión no contiene lo que el argumento necesita que contenga.

II.2. El problema de aritmética

En mi cuerpo hay del orden de **10^{13} unidades autopoieticas**. Cada una de mis células satisface la definición: produce sus componentes, realiza su membrana, mantiene su clausura de producción, es precaria, puede morir. Cada una es, con todo derecho y en el sentido estricto de la teoría, un sistema autopoietico.

Y ninguna soy yo.

Ésta es la dificultad en su forma más simple, y su simplicidad me parece una virtud. Si el sujeto está donde está la auto-producción —que es lo que (P1) y (P2) afirman conjuntamente—, entonces aquí hay diez billones de sujetos y ninguno escribe este documento, ninguno ve la pantalla, ninguno tiene la experiencia unificada que el argumento quería explicar y proteger. El criterio que debía entregarnos **un** portador de la experiencia entrega diez billones, cada uno con su propia y minúscula precariedad, y el que buscábamos no está entre ellos.

Registro la ironía: el argumento construido para **excluir** al silencio termina **multiplicando** sujetos

dentro del carbono. No da pocos sujetos; da demasiados, y ninguno el correcto.

Anticipo la respuesta inmediata: «pero el organismo también es una unidad autopoiética, de orden superior». Es la respuesta que la tradición da, y la examino en II.3 y en la Parte IV. Adelanto el resultado: para darla hay que dejar de hablar de autopoiesis en sentido estricto, y en cuanto se deja de hablar de autopoiesis en sentido estricto se pierde el argumento contra el silicio.

II.3. El cerebro no es una unidad autopoiética

La primera consecuencia, y la que hace de este documento un trabajo de filosofía de las **neurociencias** y no de metafísica general: **el cerebro no es una unidad autopoiética**.

El cerebro no produce sus componentes. Los producen sus células, cada una por su cuenta, mediante sus propias redes metabólicas. Las neuronas sintetizan sus proteínas; la glía, las suyas; el cerebro, como todo, no sintetiza nada, del mismo modo en que un bosque no fotosintetiza. Tampoco realiza su propia frontera en el sentido requerido: las meninges no son producto de una red de producción cerebral cuya clausura ellas garanticen, como la membrana sí lo es de la red metabólica celular. En la terminología de la propia tradición, el cerebro es un sistema de **segundo orden**: una unidad compuesta de unidades autopoiéticas acopladas, que **no es ella misma autopoiética en sentido estricto**.

Esto no es una objeción externa ni un hallazgo mío: es la doctrina de la teoría, correctamente enunciada. Sus consecuencias, en cambio, son severas. La posición que examino invoca la neurociencia —interocepción, correlatos neurales, regulación homeostática— para localizar la experiencia; y la neurociencia localiza los correlatos en el sistema nervioso. Pero el criterio que la posición usa para individuar sujetos **no clasifica al sistema nervioso**. La teoría localiza al sujeto (en la célula) exactamente donde la neurociencia que invoca **no** localiza los correlatos, y no lo localiza (en el cerebro) exactamente donde sí los localiza. Ese desajuste no es un detalle de implementación: es el argumento apuntando a una escala y la evidencia a otra.

II.4. El hígado es tan autopoiético como el cerebro

La segunda consecuencia es un test de discriminación. Tomemos el criterio en serio y apliquémoslo a dos órganos.

El **hígado** es un agregado de células que se autoproducen. El **cerebro** es un agregado de células

que se autoproducen. Bajo la descripción autopoietica estricta son exactamente lo mismo: sistemas de segundo orden compuestos de unidades autopoieticas de primer orden. El hígado es, de hecho, metabólicamente más activo, su recambio de componentes es mayor y su regeneración tras una lesión es espectacular donde la del cerebro es pobre. Si midiéramos «cantidad de auto-producción» por unidad de masa, ganaría el hígado.

Si la auto-producción individuara al sujeto, **nada** en el criterio diría por qué el sujeto es el cerebro y no el hígado. Y sin embargo nadie —ni Thompson, ni Di Paolo, ni Seth, ni yo— cree que el hígado sea el sujeto. La creencia común es correcta; lo que falla es el criterio, que no la respalda. Un criterio del sujeto que no distingue entre dos órganos del mismo cuerpo no es todavía un criterio del sujeto: es, a lo sumo, un criterio de lo vivo aplicado a granel.

II.5. La integración causal hace todo el trabajo, y es sustrato-neutral

La tercera consecuencia es la decisiva, y es donde el argumento se cierra.

Para escapar de II.2, II.3 y II.4 hay que **añadir algo**. Hay que añadir un principio que seleccione una escala entre las escalas anidadas —célula, tejido, órgano, sistema, organismo— y diga: el sujeto está *aquí*, y no en las de abajo ni en las de al lado. Y el candidato que la tradición ofrece, y el único que funciona, es alguna forma de **integración causal**: clausura operacional, densidad diferencial de acoplamientos, unidad funcional de la red, irreductibilidad del todo a las partes. La respuesta, en cualquiera de sus versiones, dice: la escala del sujeto es aquella en la que los procesos se enlazan en una red que se sostiene recíprocamente y cuya dinámica no es la mera suma de las dinámicas de sus partes.

Concedo que esa respuesta es buena. Lo que señalo es lo que le cuesta.

Eso añadido hace todo el trabajo. Es lo añadido, y no la auto-producción, lo que selecciona la escala. Es lo añadido lo que explica por qué el cerebro y no el hígado. Es lo añadido lo que explica por qué el organismo y no la célula. Quítese la auto-producción y déjese la integración causal: se sigue teniendo un criterio de individuación, imperfecto pero operativo. Quítese la integración causal y déjese la auto-producción: se vuelve a los diez billones de sujetos. La auto-producción no explica ningún caso que la integración causal no explique ya, y sola no explica ninguno. Eso, en la economía de una teoría, tiene un nombre: es un **epíclito**. Un término que no está haciendo trabajo explicativo y que sobrevive porque la teoría se construyó alrededor de él.

Y ahora el filo. **La integración causal es una propiedad de la organización, no del material.** Se define por cómo se encadenan los procesos, no por de qué están hechos. Es exactamente el tipo de propiedad que el funcionalista concede sin pestañear —de hecho, es lo que el funcionalista ha estado diciendo todo el tiempo—, y es el tipo de propiedad que las teorías neurobiológicas de la conciencia que el seminario revisó ofrecen (Mylopoulos, 2022): espacio de trabajo global, integración de información, recurrencia. Todas ellas caracterizan la conciencia por una **estructura causal**, y ninguna, en su formulación, menciona el carbono.

De modo que el argumento del sustrato se encuentra en esta posición: la parte que le da un sujeto es sustrato-neutral, y la parte que excluye sustratos no le da un sujeto. Tiene las dos cosas que necesita, pero nunca al mismo tiempo.

Aquí conviene traer lo que la neurociencia del curso enseña sobre dónde sí está el correlato, y no es donde está la auto-producción. Quian Quiroga, Fried y Koch (2013) hallaron neuronas que responden selectivamente a una persona, pero advierten que su lectura es esparcida y distribuida: cada concepto tendría «un conjunto de neuronas correspondientes asignadas a él», y asociar dos conceptos exige «crear unos pocos enlaces entre los grupos de células» que representan cada uno. Cada una de esas células se autoproduce igual que las demás —igual que las del hígado en II.4—; lo que varía, y lo que hace el trabajo explicativo, es con quién está conectada. La auto-producción es constante en todo el cuerpo; el correlato no. Es la misma integración causal de la que acabo de hablar, ahora vista al nivel de la neurona individual y no del órgano.

II.6. La falacia mereológica, invertida

Ahora puedo decir por qué empecé con Bennett y Hacker (2022), y por qué esto es un problema de filosofía de las neurociencias y no una curiosidad escolástica.

La falacia mereológica, en su forma estándar, consiste en atribuir **a una parte** predicados que sólo tienen sentido aplicados **al todo**: decir que el cerebro decide, o que el hemisferio izquierdo interpreta, o que una región «cree», cuando quien decide, interpreta y cree es la persona. El error no es empírico; es de gramática filosófica, y fabrica pseudoproblemas que ninguna cantidad de datos resuelve.

El argumento del sustrato comete un pariente cercano, y con la dirección invertida. Elige un criterio —la auto-producción— que **sólo se satisface en las partes**, en las células, y pretende que

ese criterio decida sobre **el todo**, sobre la persona. No dice «el cerebro decide»; dice, implícitamente, «lo que vale de la célula vale del sujeto». En ambos casos el defecto es el mismo: se predica a través de una frontera mereológica sin haber pagado el precio de justificar el tránsito. Y en ambos casos el síntoma es el mismo: una discusión que parece empírica —«¿tiene el silicio lo que hace falta?»— y que no lo es, porque su desacuerdo real está antes, en cuál es el sujeto del que se discute.

Que la frontera del sujeto no sea un dato que el mundo entrega ya hecho lo enseñó, además, el propio seminario. El grillo robot de Webb (1996) resuelve la fonotaxis delegando el cómputo a un tubo de desfase: la geometría del tubo, materia no viva y desde luego no autopoietica, es **cognitivamente constitutiva** de la solución. No es un accesorio del sistema que computa; es parte de lo que computa. Dónde termina el sistema cognitivo y empieza su entorno no es algo que se lea de la naturaleza: es algo que se decide, y la decisión tiene consecuencias.

Parte III. La clínica ya separó las variables

La Parte II mostró, por argumento, que la auto-producción no puede individuar al sujeto: en un cuerpo hay demasiadas unidades autopoieticas y ninguna es la persona. El punto no depende sólo de ese argumento. La neurología clínica ya separó, en pacientes reales y sin proponérselo, las dos variables que el argumento del sustrato funde en una sola: cuánto se autoproduce un sistema y si hay alguien ahí. No hace falta esperar una máquina consciente para poner a prueba la tesis de la Parte II; basta con leer lo que la clínica ya sabe.

III.1. El estado vegetativo: auto-producción completa, sujeto ausente

Laureys (2007) describe el estado vegetativo como la disociación de los dos componentes clásicos de la conciencia —vigilia y consciencia—: «la vigilia permanece intacta, pero la consciencia —que abarca todos los pensamientos y sentimientos— queda abolida». La descripción clínica es precisa y merece leerse con el mismo rigor con que la Parte II leyó la definición de autopoiesis. Estos pacientes tienen ciclos de sueño-vigilia, abren los ojos, respiran sin asistencia mecánica y ejecutan movimientos espontáneos; ninguno de estos signos falta. En los términos de la Parte II: la auto-producción, en el organismo entero y en cada una de sus partes, está **completa**. Las del orden de

10¹³ unidades autopoieticas que en II.2 no bastaban para dar un sujeto en un cuerpo sano siguen ahí, produciéndose, en el cuerpo del paciente. El organismo se automantiene; está vivo sin que quepa discusión alguna sobre el punto. Y, en la categoría clínica así definida, el sujeto no está.

No hay aquí ninguna premisa nueva sobre autopoiesis o silicio —Laureys no discute ninguno de los dos, y la inferencia que conecta su hallazgo con el argumento de este documento es mía, no suya—. Lo que hay es un caso en el que las variables que el argumento del sustrato trata como si fueran una sola —auto-producción y sujeto, atadas— se mueven por separado, delante de un clínico, con un diagnóstico de por medio.

III.2. El metabolismo, descalificado por quien lo mide

Si algo pudiera parecerse a una medida clínica de «cuánto se autoproduce» un cerebro, sería su metabolismo: el consumo global de energía es, de todos los correlatos disponibles, el que mejor aproxima la intuición autopoietica de un tejido ocupado en sostenerse a sí mismo. Y es exactamente ese candidato el que Laureys descarta, desde dentro de la propia disciplina que lo mide. Reporta que «voluntarios sanos y plenamente conscientes tienen valores de metabolismo cerebral global comparables a los de algunos pacientes en estado vegetativo», y concluye, sin ambigüedad: «medir los niveles globales de consumo de energía en el cerebro no puede indicar la presencia de consciencia».

El punto no es menor para el argumento de este documento. Si el candidato empírico más cercano a la auto-producción —el metabolismo, la firma más directa de un tejido que se mantiene a sí mismo— no covaría con el sujeto, entonces ninguna versión razonablemente ajustada del criterio autopoietico va a covariar mejor. La clínica no está diciendo «la autopoiesis es difícil de medir»; está diciendo que la cantidad que más se le parece **ya se midió**, y no sirve.

III.3. Lo que sí covaría: conectividad, no producción

Laureys no deja la pregunta sin respuesta. Lo que sí covaría con la presencia del sujeto es «conectividad funcional dentro de esta red frontoparietal y con centros más profundos del cerebro, en particular el tálamo». En uno de sus casos, el estímulo sensorial llegaba a la corteza y ahí se detenía: la región activa «estaba aislada y desconectada del resto del cerebro». El tejido respondía —captaba el estímulo, se activaba, metabolizaba glucosa— y aun así el sujeto no estaba, porque

esa activación no se integraba con el resto.

Esto es, en el vocabulario de la Parte II, la integración causal (II.5) leída en una neuroimagen. No es auto-producción: es la propiedad organizacional —cómo se encadenan los procesos, con quién está conectada cada región— la que hace el trabajo de seleccionar dónde está el sujeto. El tejido está vivo, se autoproduce, metaboliza y responde; lo que falta, cuando falta, es integración.

III.4. Cautela: la categoría clínica, no el paciente

Debo ser exacto sobre el alcance de este argumento, porque es fácil sobre-extenderlo hacia donde la evidencia no llega. Laureys documenta también el fenómeno inverso: conciencia encubierta detectada por neuroimagen funcional en pacientes diagnosticados, a pie de cama, como vegetativos, y un error de diagnóstico en más de un tercio de los casos. No afirmo, ni puedo afirmar con esta evidencia, que un paciente individual dado carezca de sujeto: eso es precisamente lo que el diagnóstico por neuroimagen puede estar corrigiendo. Lo que el argumento usa es la **categoría clínica** —el estado vegetativo tal como se define por sus criterios conductuales—, no el veredicto sobre ningún paciente particular. Y dentro de esa categoría, definida con independencia de cualquier tesis filosófica sobre sustratos, la disociación entre auto-producción completa y sujeto ausente es el hallazgo que la clínica reporta.

La clínica, entonces, hace por casos reales lo que la Parte II hizo por argumento: mantiene constante la auto-producción y mueve al sujeto. El factor que covaría no es la auto-producción; es organizacional.

Parte IV. La tradición ya lo sabía

IV.1. Varela 1979: de la autopoiesis a la autonomía, y por qué

Nada de lo anterior es un descubrimiento mío, y debo decir con precisión de quién es qué.

El hecho central es éste: **Varela mismo abandonó el marco estrecho, y lo abandonó por esta razón.** En *Principles of Biological Autonomy* (1979) generaliza la autopoiesis hacia la noción más amplia de **autonomía**, articulada mediante la **clausura operacional**. Un sistema es operacionalmente clausurado cuando sus procesos constituyentes se enlazan recursivamente de

modo que el resultado de cada uno alimenta las condiciones de otro dentro de la red, y la red se sostiene como red. La autopoiesis pasa a ser, en ese marco, **un caso** de autonomía: el caso en el que la clausura se realiza mediante producción material de componentes.

¿Por qué esa generalización? Precisamente por la razón que la Parte II expone: la autopoiesis **en sentido estricto es celular**, y no llega ni al organismo ni al sistema nervioso. Varela quería una teoría de la autonomía del sistema nervioso, y una teoría del organismo, y una teoría de los sistemas sociales; y la autopoiesis, tomada literalmente —producción material de los propios componentes—, no le daba ninguna de las tres. La autonomía sí, porque **ya no exige producción material de componentes**: exige sólo que los procesos formen una red que se sostiene recíprocamente. (La edición anotada de Di Paolo y Thompson, MIT Press, 2024, es un buen índice de la vigencia de esa trayectoria dentro de la propia escuela.)

Es decir: la tradición identificó el problema de la escala antes que yo, lo tomó en serio, y se movió por él. Mi aporte, si tiene alguno, es señalar lo que ese movimiento le cuesta en el debate del sustrato.

IV.2. El filo: lo organizacional es realizable en más de un medio

Porque el movimiento tiene un precio, y creo que no se ha cobrado.

La clausura operacional es una propiedad organizacional. Se define por la forma de la red: por cómo se encadenan los procesos, por la topología de las dependencias, por la recursividad del acoplamiento. No se define por la materia de los procesos. Ésa es toda su ventaja: es lo que le permite aplicarse al sistema nervioso, que no produce materialmente sus componentes, y al organismo, que tampoco.

Pero una propiedad definida por la forma de la red es, **por definición**, el tipo de cosa que puede realizarse en más de un medio. Ése es el concepto mismo de propiedad organizacional. No es que sea un accidente desafortunado que la clausura operacional resulte múltiplemente realizable: es que ser múltiplemente realizable es lo que significa ser organizacional. Y la realizabilidad múltiple es **la tesis del adversario** (Putnam, 1967; Fodor, 1974).

De ahí el dilema, que enunció como la conclusión central de este documento:

O bien se retiene la **autopoiesis estricta** —objetiva, material, de extensión celular—,

y entonces se tiene un criterio que excluye al silicio pero **no individúa al sujeto**: da diez billones de unidades, ninguna de ellas la persona, y no clasifica al cerebro.

O bien se asciende a la **autonomía** y la **clausura operacional** —y entonces se tiene un criterio que **sí individúa** organismos y sistemas nerviosos, pero que, por ser organizacional, **no excluye en principio ningún sustrato**.

No hay tercera posición en la que se tengan a la vez las dos cosas que el argumento necesita. **El camino que el enactivismo recorrió para poder individuar al sujeto es exactamente el mismo que le retira el argumento anti-funcionalista**. No es que el enactivismo se haya equivocado al recorrerlo: hizo bien, porque la autopoiesis estricta efectivamente no llegaba al organismo. Es que al llegar al organismo llegó también, sin advertirlo, al terreno del funcionalista.

Quiero marcar el estatus de esta afirmación: es un argumento conceptual, y su fuerza depende de que «propiedad organizacional» signifique lo que he dicho. Un enactivista podría responder que la clausura operacional biológica involucra esencialmente la **precariedad material**, y que ésa no es organizacional. Es la mejor respuesta disponible y la trato en VI.4; adelanto que, si se la toma, se vuelve a necesitar la auto-producción, y con ella vuelven los diez billones.

IV.3. Beer 2004: nítido en el caso paradigmático, indeterminado en los disputados

Añádase la advertencia que la teoría se hace a sí misma, y que quiero presentar con exactitud porque es fácil abusar de ella.

Beer (2004) analiza un *glider* del juego de la vida de Conway como candidato a sistema autopoietico. Es importante decir qué clase de texto es: es un **homenaje explícito a Varela**, escrito desde dentro de la tradición, y su propósito es **mejorar** la caracterización formal de la autopoiesis, no destruirla. Quien lo use como un arma anti-enactivista lo está falsificando. Yo no lo uso así: lo uso como lo que es, un testimonio de que el problema que señalo es un problema **abierto reconocido dentro de la teoría**.

Beer se detiene, antes de responder nada, en las preguntas previas:

«¿Merecen realmente los estados de las celdas individuales del retículo llamarse componentes? ¿Cuenta realmente encender o apagar una celda como producción de componentes? ¿Posee realmente un *glider* una frontera que lo genera y lo constriñe?»

Y observa dos cosas que van directamente a mi cargo:

«El hecho de que la continuidad de la identidad del *glider* dependa crucialmente de cómo elegimos definir la organización del *glider* demuestra un problema potencial.»

«Estrictamente hablando, ningún sistema es autopoietico si se lo observa durante un intervalo de tiempo suficientemente largo. La caracterización formal de la organización es un problema abierto (Fontana y Buss, 1994).»

Su conclusión es la mía, trasladada: **el criterio es nítido en el caso paradigmático —la célula— y se vuelve indeterminado justo en los casos disputados**, que son los únicos donde el argumento del sustrato lo necesita. En la célula no hay duda de qué cuenta como componente, ni de dónde está la frontera, ni de cuál es la unidad. En el *glider*, en el sistema nervioso, en el organismo compuesto, en el chip: la respuesta depende de cómo hayamos elegido describir la organización. Y el argumento del sustrato no se juega en la célula. Se juega, precisamente, en los casos disputados.

Insisto en el matiz, porque es la diferencia entre un buen argumento y uno malo: Beer no dice que la autopoiesis sea subjetiva. Dice que su caracterización formal está incompleta y que la individuación es un problema abierto. Yo no necesito más que eso.

Parte V. El laboratorio: qué mide y qué no

Cada uno de los diez experimentos de este laboratorio nace de una lectura del seminario: la jerarquía visual de Zeki (1992), las células de concepto de Quian Quiroga, Fried y Koch (2013), la neuroquímica de LeDoux (1994), la plasticidad de Hinton (1992), el grillo de Webb (1996). Cinco lecturas, diez experimentos, un solo patrón: ninguno encontró una diferencia de sustrato. Los que sobrevivieron a la auditoría que sigue devolvieron la constante que yo mismo había escrito antes de correrlos; otros ni siquiera llegaron a tener una constante que devolver. Lo consigno con el detalle que sigue porque un modelo mal construido siempre devuelve su estipulación, y eso es un hecho sobre mí, no sobre el mundo —y un patrón que se repite cinco veces, sobre cinco lecturas distintas, es un dato más interesante que cualquier resultado aislado hubiera sido. Marco además el estatus de todo lo que sigue: son **hechos verificados sobre mi propio código**, obtenidos auditándolo,

y varios de ellos son desfavorables para mí. Lo que un experimento necesitaría, para contar en contra de la Parte II, es una medida de la firma autopoiética invariante **ante el corte**: que diera igual contando o no, por ejemplo, la fuente de alimentación de la máquina como parte del sistema. Ése fue el objetivo del Experimento 10 (V.4), y es donde el laboratorio más se acerca a decir algo —aunque termine diciendo lo contrario de lo que buscaba—. El argumento de este documento no lo llevan estos experimentos: lo llevan las Partes I a IV. Lo que sigue es la auditoría, no la premisa.

V.1. Qué pretendía el laboratorio, y por qué era la pregunta equivocada

Construí un laboratorio computacional de diez experimentos con la intención de dar al argumento del sustrato una base cuantitativa: mostrar, con cifras, que hay una diferencia medible entre el modo «carbono» y el modo «silicio» de resolver un problema, y que esa diferencia favorece al carbono en dimensiones que importan.

El defecto de fondo es sencillo de enunciar y me costó verlo: **no hay carbono en el laboratorio**. Los diez experimentos corren en silicio. Todos. Lo que comparan no es carbono contra silicio: comparan **pares de modelos**, ambos ejecutados en la misma máquina, y **bautizan** a uno de cada par «carbono». El Experimento 5 compara STDP contra retropropagación: son **dos algoritmos**. El Experimento 1 compara una red densa contra una convolucional: son dos arquitecturas, y la ventaja de la convolucional es una ventaja **explotada en silicio desde Fukushima (1980) y LeCun (1998)**, es decir, un logro de la ingeniería de silicio que yo estaba atribuyendo al carbono por vía de etiqueta. En cada caso, **la etiqueta hace el trabajo**. El laboratorio no medía un sustrato; medía dos modelos y le ponía a uno el nombre del sustrato que yo quería que ganara.

Retrospectivamente, esto era predecible desde la Parte II: si la individuación de la unidad es previa a la medición, entonces ninguna medición puede entregarla. Yo estaba pidiéndole al laboratorio que decidiera dónde estaba el sujeto, y el laboratorio sólo podía devolverme el recorte que yo había codificado en él.

V.2. Clasificación epistémica honesta

Auditando las once cifras publicadas, **ocho son consecuencias analíticas de constantes que yo mismo elegí**. No son hallazgos: son teoremas triviales sobre mis propios parámetros, presentados con la tipografía de un resultado empírico. Tres ejemplos, para que se vea la forma del error:

- El «90 %» del Experimento 1 es $1 - P_{\text{CONEXION}}$. Es decir: fijé una probabilidad de conexión y después «descubrí» su complemento.
- El «80,0 % frente a 1,03 %» del Experimento 2 es el valor esperado del solapamiento de vectores Bernoulli(p): $E[\text{solapamiento}] = p$. Elegí p y reporté p .
- El «512×» del Experimento 5 se sigue de la localidad de la regla de actualización, y **no depende de N**: no es una medida de escalamiento, aunque estaba presentado como si lo fuera.

Un cálculo cerrado puede ser informativo —puede hacer vívida una consecuencia no obvia—, pero **no es evidencia** de nada sobre el mundo, y presentarlo con el formato de un experimento es engañoso aunque quien lo haga no tenga intención de engañar. Yo no la tenía; el efecto es el mismo.

La única medición **empíricamente genuina** del laboratorio es el benchmark escalonado de consumo energético, y es la que más me enseñó, porque fue ahí donde empecé buscando la prueba de que el silicio es energéticamente inviable.

El benchmark recorre diecinueve configuraciones —de CPU a una sola GPU, de una sola GPU a dos en paralelo, y de ahí a un régimen híbrido CPU+GPU+GPU, cada tramo con su propio rango de tamaño de problema—. La brecha de eficiencia entre carbono y silicio, medida por evento sináptico, va de 2.855 a 409.761 ($2,9 \times 10^3$ a $4,1 \times 10^5$) a lo largo de esas diecinueve configuraciones —se mueve ciento cuarenta veces (143×)— y no lo hace monotónicamente: el mínimo no está en ningún extremo de la tabla, sino en medio, y repartir el mismo cómputo entre dos tarjetas en vez de una, lejos de mejorar la eficiencia, 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, con pacientes y no con tarjetas gráficas, ya había llegado a la misma advertencia desde la clínica: «medir los niveles globales de consumo de energía en el cerebro no puede indicar la presencia de consciencia» (Laureys, 2007). Mi intuición de partida no era falsa; estaba mal dirigida. Medía con cuidado la magnitud equivocada.

Y aquí quiero reconocerle algo al texto original: ya la descontaba, describiéndola como «una cota superior de una comparación deliberadamente desfavorable». Es decir: la única cifra real del laboratorio es una cota **amañada contra el silicio**, y el texto lo sabía. Sostengo esa lectura y la refuerzo.

V.3. Los datos fabricados del Experimento 1

Debo decir esto explícitamente, y no en una nota al pie.

El Experimento 1 publicaba columnas de tiempo y energía que eran **literales escritos a mano**, comentados en el código como «tiempo medio empírico», con un cronómetro declarado en el mismo archivo y **nunca leído**. No eran mediciones. Eran números que yo había escrito y presentado como si un instrumento los hubiera producido.

No hubo intención de fraude: fueron, en su origen, valores de relleno para probar el pipeline de graficación, que sobrevivieron a la limpieza y se publicaron. Pero la intención no cambia el estatus de la cifra, y una explicación no es una excusa. **Ya está corregido**: esas columnas se eliminaron. Lo consigno aquí porque un trabajo que denuncia el autoengaño ajeno y esconde el propio no vale nada, y porque el hecho de que yo mismo haya podido publicar un número fabricado sin advertirlo es, precisamente, un dato sobre lo fácil que es que la etiqueta preceda a la medición.

V.4. El Experimento 10 y κ : la historia completa

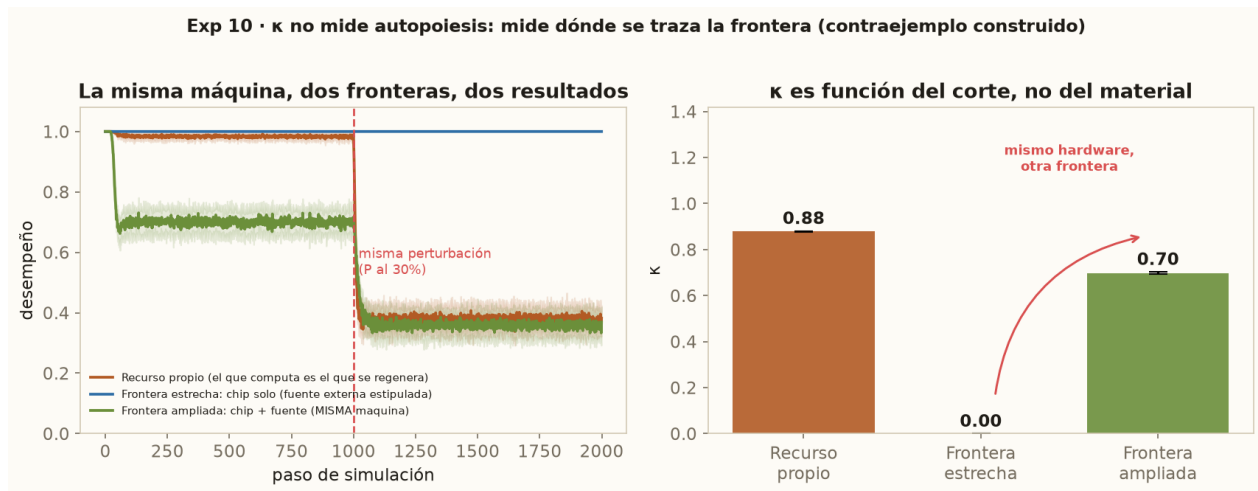


Figura 1: *Experimento 10 (reformado)*. La misma máquina, bajo la misma perturbación, con dos fronteras distintas: contando sólo el chip, $\kappa = 0,00$ (inmune); contando también su fuente de alimentación, $\kappa = 0,70$ (24 semillas). El sustrato no cambió: cambió el corte. κ mide compartición de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

El Experimento 10 era la joya, y es la pieza que más aprendí.

Modelé dos sistemas que computan bajo una «lesión metabólica» y definí un coeficiente κ de acoplamiento entre el cómputo y el automantenimiento. Lo presenté como **la firma operacional**

de la autopoiesis: $\kappa \approx 0,88$ para el sistema acoplado («carbono»), $\kappa = 0,00$ para el desacoplado («silicio»). La lectura era: aquí está, medida, la diferencia.

La auditoría mostró tres cosas.

Primera: $\kappa = 0,00$ no es un resultado. En esa rama el recurso se repone **por estipulación** en cada paso, y el parámetro lesionado **no aparece en la ecuación**. Envenenarlo con valores absurdos sigue devolviendo cero. Es una identidad aritmética disfrazada de hallazgo: el cero no lo produjo el silicio, lo produjo mi estipulación de que la fuente era inagotable.

Segunda: $\kappa \approx 0,88$ tampoco es un descubrimiento. Tiene forma cerrada y recorre **todo el rango [0,1]** moviendo un solo parámetro. No es una constante de la naturaleza; es una perilla.

Tercera, y la que importa: lo que κ mide, cuando mide algo, es **compartición de un presupuesto de recursos**. Nada más. Un portátil cuyo cómputo compite con su refrigeración por el mismo presupuesto eléctrico —cualquier portátil con *throttling* térmico— tiene $\kappa > 0$. Si κ fuera la firma de la autopoiesis, mi computador estaría vivo.

Y peor, en el sentido que más me interesa: **ninguna de mis dos ramas es autopoietica**. En ambas, la bomba homeostática es un parámetro **exógeno** que yo lesione desde fuera. No hay clausura de producción: el sistema no produce los componentes que lo producen. No hay adaptividad: el sistema no regula activamente su posición respecto de sus condiciones de viabilidad, que es justo lo que Di Paolo (2005) exige para que haya normatividad. Construí un modelo para medir la autopoiesis y ninguno de sus dos brazos era autopoietico. El modelo no podía ni siquiera formular su objeto.

El resultado nuevo. Y aquí está la comprobación que cierra el asunto, ejecutada con 24 semillas y con el mismo hardware:

Rama	Corte	κ
Recurso propio	el que computa es el que se regenera	$0,88 \pm 0,00$
Frontera estrecha	chip solo, fuente externa estipulada	$0,00 \pm 0,00$
Frontera ampliada	chip + fuente , la misma máquina	$0,70 \pm 0,01$

Léase la tercera fila contra la segunda. Es **el mismo silicio**, bajo **la misma perturbación**, con **el mismo código**. Lo único que cambió es **dónde tracé la frontera del sistema**: si cuento la fuente de alimentación como parte del sistema en vez de como un exterior inagotable estipulado. Y κ pasa de 0,00 a 0,70 — del orden del 0,88 del «carbono».

No cambió el sustrato. **Cambió el corte.**

La conclusión que extraigo es exactamente la de la Parte II, ahora en forma de resultado: κ **no mide una propiedad del sustrato; mide dónde tracé la frontera**. La medida no descubre unidades: las **presupone**. Y no podía ser de otro modo, porque toda medida necesita un sistema al que aplicarse, y decidir cuál es ese sistema es la tarea previa que ninguna medición puede hacer por sí misma. Construí la operacionalización que debía darle al argumento del sustrato su base objetiva, la vi fallar, y el diagnóstico de por qué **tenía** que fallar es el argumento de este documento. Considero este resultado negativo el aporte más sólido del trabajo, precisamente porque no lo buscaba.

V.5. Qué se conserva de los cinco ejes

No propongo tirar el laboratorio. Propongo **reencuadrarlo**, y creo que reencuadrado sigue siendo útil.

Los cinco ejes que los experimentos exploran —esparsidad y localidad de la conexión; solapamiento y capacidad representacional; localidad de la regla de aprendizaje; acoplamiento entre cómputo y mantenimiento; costo energético por operación— **no miden «carbono contra silicio»**, y esa lectura debe abandonarse sin residuo. Lo que sí hacen, y hacen legítimamente, es comparar **esquemas organizacionales**: dos maneras distintas de organizar un sistema que computa, con sus costos y sus ventajas, ambas implementables en cualquier sustrato que sostenga las operaciones.

Bajo ese reencuadre, los experimentos dejan de ser evidencia sobre el sustrato y pasan a ser **ilustraciones de tesis organizacionales**, que es lo que siempre fueron. Y ese reencuadre no es una retirada: es la posición correcta, y es coherente con el argumento entero. Si la Parte II tiene razón y lo que hace el trabajo es la organización y no el material, entonces un laboratorio que compara organizaciones está midiendo exactamente lo que hay que medir. Lo único que estaba mal era el nombre que le puse a cada rama.

Lo que sí sobrevive como distinción sustantiva, y lo ofrezco como **reencuadre y no como**

criterio cerrado, es la diferencia entre automantenimiento **delegado** y **constitutivo**: entre un sistema cuyas condiciones de persistencia son internas a la unidad y uno que las tiene fuera, en técnicos y redes eléctricas cuya normatividad es previa y ajena. Esa diferencia es real y es la intuición sana que había en todo esto. Pero —y ésta es la lección del Experimento 10— es **relativa al corte**. Y decidir el corte es la tarea que el debate ha estado saltándose.

Parte VI. Objeciones y límites

VI.1. «Sigue siendo necesaria»: condición versus criterio de individuación

La objeción más fuerte que se le puede hacer a este documento no niega nada de lo mostrado: lo concede y pregunta si basta. Dice así: «de acuerdo, la auto-producción no individúa al sujeto; la Parte II lo mostró y la Parte III lo confirmó en la clínica. Pero sigue siendo **necesaria**. El silicio no se autoproduce, luego le falta una condición sin la cual no hay conciencia, individúe o no».

Concedo la premisa entera y niego que salve la conclusión que el argumento del sustrato necesita. Este documento nunca ha negado que la vida sea condición necesaria de la conciencia: lo dije en la introducción y lo repito aquí, porque es el punto en el que más fácil es que se me malentienda. Lo que niego es que la auto-producción sea el **criterio que individúa al sujeto** —el que decide, entre los sistemas anidados en un cuerpo, cuál de ellos siente—. Una condición necesaria y un criterio de individuación son dos cosas distintas, y el argumento del sustrato necesita las dos a la vez para llegar a su conclusión sobre el silicio; sin la segunda, de «el silicio no se autoproduce» no se sigue nada sobre lo que el silicio pueda o no sentir, aunque la primera sea verdadera.

Para verlo, hay que preguntar algo que la objeción da por resuelto: ¿de qué, exactamente, es necesaria la autopoiesis?

Si es necesaria del sujeto —si lo que se afirma es que el portador de la experiencia tiene que ser, él mismo, una unidad autopoietica—, entonces la objeción hereda intacto el problema de la Parte II. El sujeto tendría que ser una unidad autopoietica, y ninguno lo es: somos, como mostró II.3, sistemas de segundo orden, compuestos de unidades autopoieticas sin ser ellos mismos autopoieticos en sentido estricto. Quien exija auto-producción al portador de la experiencia se queda sin portador también en el carbono, por la misma aritmética de II.2. La objeción, en esta lectura, no rescata el

argumento: lo hunde más rápido, porque convierte en necesaria una propiedad que ni el cerebro humano posee.

Si es necesaria de los componentes —si lo que se afirma es, más modestamente, que el material del que está hecho el sujeto tiene que ser capaz de autoproducirse, aunque el sujeto mismo no lo sea—, entonces la objeción ha cambiado de tesis sin decirlo. Ya no es una tesis sobre el sujeto, sino sobre el sustrato en que la organización que sí individúa (la integración causal de II.5) se realiza. Y esa tesis necesita un argumento adicional, uno que diga por qué ese material y no otro es apto para soportar la organización relevante. La autopoiesis no puede dar ese argumento, porque la autopoiesis no habla de sujetos ni de organizaciones que individúan: habla de **células**, de si una red metabólica produce su propia membrana. Que las células se autoproduzcan no dice nada, por sí solo, sobre si el silicio puede o no sostener la clase de integración causal que la Parte II identificó como lo que hace el trabajo. Para eso hace falta otro argumento, y éste no lo es.

Una condición de fondo, en suma, no es un criterio de exclusión. Que todo lo consciente conocido resulte ser vivo —si lo es— no implica que sólo lo vivo pueda serlo; implica, a lo sumo, que hasta ahora no hemos visto el caso que lo decidiría. El argumento del sustrato quiere extraer una prohibición metafísica de una correlación conocida, y ese paso es el que la distinción entre condición y criterio bloquea. Esto es lo que blinda a este documento frente al arma más fácil que se le podía apuntar: no porque yo niegue que la vida importe, sino porque nunca necesité negarlo. Bastaba con que importar no fuera lo mismo que individuar.

VI.2. «Caricaturizas a Thompson y a Di Paolo»

Es la siguiente objeción seria, y la que un evaluador competente hará en cuanto conceda VI.1. Dice: ni Thompson ni Di Paolo sostienen jamás que la célula sea el sujeto de la experiencia. Hablan de **autonomía del organismo**, de sistemas de identidad de orden superior, de agencia adaptativa. Tu problema de aritmética ataca una posición que nadie defiende.

Concedo el hecho, y niego que me perjudique: es mi argumento.

En efecto, no sostienen que la célula sea el sujeto. Precisamente por eso pasan a la autonomía. Y ese paso es lo que estoy documentando, no lo que estoy negando. Mi tesis no es «los enactivistas creen que la célula siente». Mi tesis es: **en cuanto la tradición pasa a la autonomía para poder individuar al organismo, deja de tener un argumento *material* contra el silicio y pasa a tener**

una teoría organizacional de la unidad, compatible con la realizabilidad múltiple. La concesión que la objeción me pide es la premisa que necesito. Cuanto más se insista en que Thompson y Di Paolo hablan de autonomía y no de autopoiesis celular, más completo queda el cuerno organizacional del dilema de IV.2.

No sostengo que se equivoquen sobre la vida. Creo que su teoría de la vida es la mejor que hay. Sostengo algo más limitado: que su mejor teoría de la individuación **no puede hacer el trabajo anti-funcionalista** que se le encarga en el debate del sustrato, y que el encargo se le hizo sin advertir el precio.

VI.3. «Convencionalidad no es relatividad al observador»

La objeción dice: estás deslizándote de «la definición de organización tiene un momento convencional» a «la autopoiesis es relativa al observador», y ese deslizamiento es un error categorial. Fijada una granularidad, si una bacteria se autoproduce es un hecho físico. Tu uso de Beer repite el error de decir «la vida es ambigua en los virus, luego la vida no existe».

Concedido, íntegramente. Y quiero registrar que una versión anterior de mi propio argumento cometía exactamente ese error, en forma de *tu quoque* contra Searle: «la adscripción de autopoiesis es relativa al observador, igual que la sintaxis; luego no funda nada». La descarté por tres razones, todas buenas: (i) el error categorial que la objeción señala; (ii) el *tu quoque* no es simétrico, porque la sintaxis es más radicalmente no-física que la auto-producción, que es una propiedad química; (iii) malinterpretaba a Beer, cuyo texto es un homenaje que plantea un problema abierto para mejorar la teoría.

El argumento que defiende **no dice** que la autopoiesis sea subjetiva. Dice que es **objetiva y de extensión equivocada**. Es una tesis sobre **extensión**, no sobre relatividad al observador, y la diferencia no es cosmética: mi cargo sobrevive intacto aunque la autopoiesis sea la propiedad más objetiva del universo. De hecho lo necesito objetiva: si fuera vaga, mi problema de aritmética no tendría fuerza, porque no habría un hecho determinado sobre cuántas unidades autopoieticas hay en mi cuerpo. Lo hay, y son demasiadas.

VI.4. «¿Y si la integración causal tampoco individúa?»

La objeción más interesante, y la que un enactivista sofisticado debería hacer, es ésta: tu escape —la integración causal— sufre del mismo mal. También la integración causal depende de cómo se describa el sistema; también ella admite cortes múltiples; también ella, con otra granularidad, da otra respuesta. No has resuelto el problema de la individuación: lo has mudado.

Concedo el antecedente, y sostengo que la consecuencia me favorece.

Si la integración causal tampoco individúa, entonces el problema de la individuación es **general**, previo, y no está resuelto por nadie. Y entonces mi conclusión no se debilita: se refuerza. Porque lo que sostengo no es «la integración causal individúa y la autopoiesis no». Sostengo que **el argumento del sustrato necesita una individuación que no tiene**, y esta objeción dice que nadie la tiene. En ese escenario, el argumento del sustrato está peor, no mejor: pretendía derivar una conclusión ontológica fuerte —el silicio no siente— de un criterio de individuación, y resulta que no hay criterio de individuación disponible para nadie. De un problema abierto para todos no se sigue una exclusión para uno.

Una variante enactivista mejor responde: la clausura operacional biológica no es puramente organizacional, porque involucra esencialmente la **precariedad material** —el sistema se desintegra si la red se detiene—, y eso no es forma, es materia. Es la mejor respuesta disponible y la tomo en serio. Pero obsérvese lo que hace: reintroduce la exigencia material para bloquear la realizabilidad múltiple, y al reintroducirla vuelve a caer en el primer cuerno del dilema. Si la precariedad material es esencial a la clausura, entonces la clausura vuelve a satisfacerse allí donde hay producción material de componentes —en las células— y volvemos a los diez billones. El dilema no es un accidente retórico: es una tijera, y ambos cortes están afilados.

VI.5. El zombi autopoietico

Un límite que debo reconocer, y que en parte me impuso el examen de una versión anterior de este trabajo.

Supóngase un sistema plenamente autopoietico, adaptativo, precario, operacionalmente clausurado, funcionalmente indistinguible de un organismo, y sin experiencia alguna. ¿Es concebible? Me parece que sí, y me parece que es tan concebible como el zombi de silicio. La

autopoiesis, tal como la teoría la caracteriza, es una **descripción estructural y organizacional**; y de una descripción estructural no se sigue, sin más, la presencia de experiencia fenoménica. Quien crea que el zombi funcional de silicio es concebible debería, por paridad de razones, conceder el zombi autopoietico.

Registro esto porque es un límite y no una victoria. No lo uso como argumento contra el enactivismo —sería un mal argumento, del tipo que reprocho a otros—: lo uso para marcar que **mi tesis no necesita resolver el problema difícil, y no lo resuelve**. Y también porque, si el zombi autopoietico es concebible, entonces la auto-producción no sólo falla en individuar: tampoco garantiza lo que quería garantizar. Pero esa es una consideración adicional, más débil que la principal, y no apoyo nada en ella.

VI.6. Antecedentes: qué es mío y qué no

No finjo prioridad, y quiero ser explícito sobre las deudas.

La distinción entre identidad **propia** y **prestada** —la del artefacto, cuya identidad «le es acordada» por otro— es de **Jonas (1968)**, en «Biological Foundations of Individuality». Toda la intuición que anima (P2) está ahí, y mejor dicha.

Que el problema de la individualidad sea «el de justificar cuál separación elegimos entre el gran conjunto de distinciones posibles y arbitrarias» está dicho, **con esas palabras**, en **Barandiaran, Di Paolo y Rohde (2009)**. Es decir: el problema que constituye el núcleo de mi cargo fue enunciado con toda claridad, hace más de quince años, por dos de los autores de la tradición que critico, uno de ellos Di Paolo. No lo descubrí.

Beer (2004) planteó la individuación como problema abierto **dentro** de la teoría autopoietica, en un homenaje a Varela. **Varela (1979)** hizo la generalización a la autonomía. **Di Paolo (2005)** mostró que la autopoiesis desnuda no basta.

¿Qué queda que sea mío? Tres cosas, y las acoto:

1. **La aplicación del problema de la escala al debate de la conciencia artificial**, con el giro de que la trayectoria del propio Varela (autopoiesis → autonomía, 1979) es **autosocavante** para el uso anti-funcionalista de la autopoiesis. No descubro la condición de individualidad: **la cobro donde no se la había cobrado**.

2. **La demostración construida**, como resultado negativo: construí κ como firma operacional de la autopoiesis, la auditué, y medía compartición de presupuesto de recursos y, en el fondo, mi propio corte.
3. **El remanente utilizable**: la distinción automantenimiento delegado vs. constitutivo, y la constatación de que es relativa a la frontera. Ofrecida como reencuadre, no como criterio.

Formulado ante un evaluador hostil, en una oración: *no reclamo haber descubierto que la individuación es un problema para la autopoiesis —Beer lo planteó desde dentro de la teoría—, sino haber mostrado que ese problema abierto es **fatal precisamente para el uso que la tradición hace de la autopoiesis en el debate del sustrato**, y haberlo mostrado construyendo la métrica que debía evitarlo y viéndola medir mi propia estipulación.*

VI.7. Dónde se detiene el argumento

Marco los límites con precisión, porque una tesis que no dice dónde para no dice nada.

No he mostrado que una máquina de silicio sienta. No tengo la menor idea de si puede, y este documento no aporta ninguna razón para creer que sí. Mostrar que un argumento contra P falla no es mostrar P.

No he mostrado que la vida sea irrelevante para la mente. Puede ser perfectamente cierto que sólo los sistemas vivos sean conscientes. El argumento de Seth (2025) no depende de la premisa que ataco: su eje es vivo/no-vivo y él mismo rechaza el chauvinismo del carbono, y nada de lo que digo lo toca.

No he explicado los qualia. Mi argumento es **neutral ante el problema difícil**, y la neutralidad es deliberada, no una omisión: vale sea el funcionalismo verdadero o falso. Es **compatible con la realizabilidad múltiple** sin implicarla.

No he refutado el enactivismo. He sostenido que una de sus aplicaciones —la anti-funcionalista— se apoya en una unidad que no tiene.

Lo mostrado es más modesto y, creo, más firme: **una de las dos grandes objeciones al funcionalismo se apoya en una unidad que no tiene.**

Conclusión

El argumento del sustrato parecía fuerte porque parecía tener dos cosas a la vez: un criterio objetivo y material —la auto-producción— y un sujeto al cual aplicarlo. Tiene sólo una por vez.

En sentido estricto, la autopoiesis es objetiva y señala **células**: diez billones de unidades en un cuerpo, ninguna de ellas yo, y un cerebro que no clasifica porque no produce sus componentes, y un hígado que clasifica exactamente igual que el cerebro. En el sentido amplio al que la tradición se vio **obligada** a migrar —autonomía, clausura operacional, Varela 1979—, individúa organismos y sistemas nerviosos, que es lo que hacía falta; pero se ha vuelto una propiedad de la **organización**, y las propiedades de la organización no excluyen sustratos: son la tesis del adversario. El camino que el enactivismo recorrió para individuar al sujeto es el mismo que le retiró el argumento contra el silicio. La clínica lo dice a su manera, sin filosofía de por medio: donde la auto-producción está intacta y el sujeto no está, lo que decide es la conectividad.

Bennett y Hacker enseñaron que atribuir a la parte lo que es del todo no es un desliz de estilo sino un error que fabrica pseudoproblemas. El argumento del sustrato es un caso, con la dirección invertida: elige un criterio que sólo se satisface en las partes y pretende que decida sobre el todo. Discute de qué está hecho el sujeto antes de poder decir cuál es.

Queda algo, y es lo que me llevo. La diferencia real no es carbono contra silicio, sino automantenimiento **delegado** contra **constitutivo**: la hay, y es genuina, entre un sistema cuyas condiciones de persistencia son internas a la unidad y uno que las tiene fuera, en técnicos y redes eléctricas cuya normatividad es previa y ajena. Ahí sí importa la energía: no por cuánta se gasta, sino por quién responde de ella. Pero esa diferencia, como me enseñó mi propio experimento —misma máquina, misma perturbación, κ de 0,00 a 0,70 al mover sólo la frontera—, es **relativa al corte**. Y decidir el corte 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.

Glosario

Una línea por término. El propósito es **distinguir sin equiparar**: buena parte del atractivo del argumento del sustrato viene de deslizarse entre estas nociones como si fueran una sola.

Término	Distinción mínima
Dependencia ambiental	El sistema requiere insumos del entorno para persistir. La tiene todo sistema físico, vivo o no; no distingue nada por sí sola.
Regulación autónoma	El sistema modula sus propios parámetros internos según su estado, sin control externo momento a momento; un termostato la tiene en grado mínimo.
Automantenimiento	El sistema actúa de modo que su propia persistencia se conserva; nada dice todavía sobre quién produce los componentes ni sobre dónde está la frontera.
Autorreparación	El sistema restaura componentes dañados a partir de un plan o capacidad ya presente; no implica que haya producido esos componentes en primer lugar.
Autoproducción	El sistema fabrica los componentes que lo constituyen; es más fuerte que el automantenimiento y que la autorreparación.
Autopoiesis	Autoproducción clausurada más producción de la propia frontera: la red produce los componentes que producen la red, y la frontera que la encierra. Objetiva; extensión paradigmática: la célula.

Término	Distinción mínima
Clausura operacional	Los procesos del sistema se enlazan recursivamente de modo que la red se sostiene como red. Organizacional: no exige producción material de componentes (Varela, 1979).
Clausura causal	Ninguna causa externa al sistema es necesaria para explicar sus transiciones internas; noción más fuerte y raramente satisfecha, no equivalente a la anterior.
Adaptividad	Capacidad de regular activamente la propia posición respecto de las condiciones de viabilidad, con grados y anticipación; añadido que Di Paolo (2005) exige sobre la autopoiesis desnuda.
Precariedad	Los procesos del sistema se desintegran si la red que los sostiene se detiene: la existencia es una tarea que puede fracasar.
Normatividad intrínseca	Hay un hecho sobre qué es mejor o peor para el sistema mismo , no relativo a un diseñador; requiere precariedad más adaptividad, no autopoiesis sola.
Agencia	Capacidad de iniciar acción según normas propias que modula el acoplamiento con el entorno (Barandiaran, Di Paolo y Rohde, 2009); presupone individualidad y normatividad, no las entrega.

Término	Distinción mínima
Cognición	Regulación del acoplamiento con el entorno sensible al significado para el sistema; en la tradición enactiva, coextensiva con la vida (Thompson, 2007) — tesis, no definición neutral.
Inteligencia	Eficacia en resolver problemas dado un objetivo; no implica experiencia ni normatividad propia, y su presencia no es evidencia de conciencia.
Conciencia	Que haya algo que es ser ese sistema. No se sigue de ninguno de los términos anteriores; el punto de este documento no es explicarla.
Unidad fenomenológica	Que las experiencias de un sujeto se den como experiencias de un solo sujeto; el explanandum que (P1) presupone y que la autopoiesis no entrega.
Condición necesaria	Un requisito sin el cual X no se da; no dice, por sí sola, cuál de varios candidatos anidados es X. Confundirla con el criterio de individuación es el error que VI.1 diagnostica en (P2).
Individuación	La determinación de cuál sistema es el sistema: dónde va el corte entre unidad y entorno. Problema previo a toda medición, y el que este documento pone en el centro.
Autonomía	Identidad sostenida por clausura operacional, con la autopoiesis como un caso (Varela, 1979). Individúa organismos; por organizacional, no excluye sustratos. Ahí está el dilema.

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.
- Quian Quiroga, R., Fried, I., & Koch, C. (2013). *Brain Cells for Grandmother*. *Scientific American*, 308(2), 30-35.
- 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.
- Seth, A. K. (2021). *Being You: A New Science of Consciousness*. Faber & Faber.
- 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.