

Desde la era cibernética hacia la nueva ciencia cognitiva ¹

DR. TOM FROESE

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
Universidad Nacional Autónoma de México

E-mail: t.froese@unam.mx
Web: <http://froese.wordpress.com>

Resumen

Wiener fundó la cibernética en el principio de auto-gobierno, por ejemplo el uso de bucles de retroalimentación negativa para el control de máquinas. Años antes, el cibernético mexicano Rosenblueth con sus colegas Wiener y Bigelow argumentaban que comportamiento dirigido a metas puede ser explicado por la retroalimentación negativa. Esta propuesta revolucionaria implicaba que nuestra experiencia al actuar intencionadamente podía hacerse compatible con una visión del mundo estrictamente científica, que sostiene que la naturaleza física no sigue ningún propósito. Sin embargo, los seres vivos no sólo son auto-gobernantes sino también, a través del metabolismo, individuos físicamente auto-productivos. Esto es de importancia para el surgimiento de una nueva ciencia cognitiva, que fundamenta el sentido de la existencia en la corporización biológica, y por lo tanto en la mortalidad.

Introducción

En el año 1943, Arturo Rosenblueth, Norbert Wiener y Julian Bigelow publicaron uno de los artículos más famosos de la época [1]. Su contribución sirvió como parte de la fundación de un agrupamiento increíblemente transdisciplinario que desde 1948 y en adelante se va a denominar cibernética, siguiendo la propuesta hecha por Wiener. El artículo trató de un análisis teórico del comportamiento orientado a una meta, cómo se manifiesta en todos los seres vivos, y también en algunos dispositivos diseñados para actuar automáticamente.

La gran innovación fue acercar la ciencia, con sus métodos y teorías objetivos, al estudio del fenómeno psicológico y subjetivo de seguir objetivos. En general, las metas son asociadas a condiciones de satisfacción y la posibilidad de fracaso, algo que no aplica a fenómenos del dominio de la física pura, por ejemplo, la trayectoria de un planeta. Basándose en la aplicación de la teoría sistémica de la retroalimentación negativa, Rosenblueth y sus colegas iban más allá del reduccionismo clásico para fundar el estudio de la mente como tal. Este fue un paso importante hacia superar el gran reto de investigar la relación mente-cuerpo.

¹ Borrador final de un artículo que va a ser publicado en 2016 en *Ciencia*, la revista de la Academia Mexicana de Ciencias, como parte de un número especial dedicado al cibernético Norbert Wiener.

Sin embargo, en retrospectiva la teoría sobre el papel de la retroalimentación negativa era demasiado general. No hay duda de que podemos experimentar en nosotros mismos el fenómeno de tener metas propias y de preocuparse por que sean exitosas. Pero ¿qué tal un termostato diseñado para mantener la temperatura por medio de la retroalimentación negativa? ¿Es correcto decir que el termostato también tiene fines propios, o bien fines *intrínsecos* como dijeron los cibernéticos Rosenblueth y colegas? Algunos investigadores en el campo de la inteligencia artificial contemporánea siguen estando de acuerdo en que no hay ninguna diferencia esencial entre un termostato y un ser vivo (incluyendo a los seres humanos). Sin embargo, un movimiento importante en la nueva ciencia cognitiva, a veces llamado el “enactivismo”, siguiendo la propuesta por el biólogo chileno Varela y sus colegas [2], propone que existen diferencias fundamentales que separan la organización de los seres vivos de la de los sistemas que son principalmente definidos por la retroalimentación negativa.

Los biólogos chilenos Maturana y Varela han destacado desde los setentas que los seres vivos deben estar definidos como seres *autopoieticos* (su termino nuevo de “autopoiesis” se traduce como auto-producción) [3]. Además, según el nuevo enactivismo, a través de la necesidad de constantemente crear y mantener su organización física, mientras que evita la desintegración y la muerte, un organismo también da a luz una perspectiva de preocupación sobre su ambiente, concertándole en un mundo significativo. En ese sentido, la nueva ciencia cognitiva argumenta que solamente los seres vivos tienen fines propios en una manera intrínseca y subjetiva. La implicación es que somos más que nuestras maquinas porque somos individuos mortales [4].

De la retroalimentación negativa a la retroalimentación doble

La retroalimentación negativa puede usarse para corregir errores, lo que es útil si queremos crear sistemas auto-reguladores. Consideremos el humilde termostato de nuevo. Por ejemplo, imagine que la llegada del verano causa un incremento en la temperatura de la casa, que es medida por el termómetro. El termostato compara esta medición con el valor deseado para el cual ha sido programado. En este caso la medida va a ser muy alta, y así, un regulador contrarresta esta divergencia; por ejemplo, incrementando la salida del sistema de enfriamiento. La influencia del calor externo es por lo tanto reducida de forma automática.

Desde la perspectiva de aquellos que viven dentro de la casa, este sistema se está comportando intencionadamente y siguiendo una meta: mantener la temperatura alrededor de cierto punto establecido. Pero, ¿Comparte el sistema este punto de vista sobre sus operaciones? Una manera de revelar el hecho de que al sistema no le importa en realidad, es exponiéndolo a una fuerte perturbación, que tiene el efecto de frustrar su “meta” y ver cómo reacciona.

Imagine que abrimos la caja con el sistema de control del clima en la casa y cambiamos un par de cables alrededor, de tal manera que el sistema envía exactamente la señal opuesta al sistema de enfriamiento. El siguiente día, cuando el calor en la casa empieza a volverse demasiado de nuevo, el termostato otra vez detecta una divergencia del valor deseado y activa el regulador. Pero esto va ahora a tener el efecto de reducir, en lugar de incrementar, la salida del sistema de enfriamiento, incrementando así aún más la temperatura. El termostato va por lo tanto a enviar una respuesta aún mayor al regulador, y esto va a disminuir todavía más el sistema de enfriamiento; y así sucesivamente. Podemos imaginar que el bucle va a continuar hasta que el sistema eventualmente se sobrecaliente y se quiebre.

La situación es muy diferente en el caso de los organismos. Nuestra temperatura es una variable esencial que el cuerpo debe regular para mantener ciertos estados fisiológicos. Por ejemplo, la gente en la casa que tuvo que enfrentar las consecuencias del sistema reconfigurado debe tener calor. Eventualmente, esto debe de haber desencadenado un caso biológico de retroalimentación negativa. Una vez que nuestro cuerpo se calienta demasiado, empieza a sudar, y la evaporación del sudor tiene un efecto de enfriamiento en nuestra piel. Pero, ¿Qué pasa cuando esta retroalimentación negativa no sirve? Si el cuerpo se sobrecalienta va a llegar un punto en el que se quiebre, colapsemos e incluso muramos. Sin embargo, antes de este momento final vamos a empezar a sentir malestar y vamos a tratar de reducir el impacto del calor de otras formas. Podemos abrir ventanas, cerrar el sistema de control de la temperatura que funciona mal, o simplemente irnos de la casa. Para explicar esta notable capacidad indefinida de adaptación, necesitamos más que retroalimentación negativa.

El psiquiatra y cibernético inglés W. Ross Ashby (figura 1) dedicó una gran parte de su vida profesional a explicar cómo es posible que los animales puedan adaptar su conducta a circunstancias imprevistas. Cuando el comportamiento no es efectivo para contrarrestar una perturbación, quería que la organización interna pudiera ser ajustada automáticamente hasta que un nuevo comportamiento efectivo fuera hallado otra vez. En lugar de mera estabilidad, Ashby estaba interesado en lo que llamó “ultraestabilidad” [5].



FIGURA 1. Dr. W. Ross Ashby
(Foto tomado de http://en.wikipedia.org/wiki/William_Ross_Ashby)

Logró célebremente diseñar y construir un sistema mecánico, hecho de los materiales excedentes de la segunda guerra mundial, que podía adaptarse espontáneamente a su organización interna de modo que sus variables esenciales volvieran a estabilizarse cuando estas excedían sus límites. Este aparato innovador es conocido como “homeostato”. No necesitamos entrar en los detalles de su funcionamiento aquí. Es suficiente que tengamos una comprensión general del concepto de ultraestabilidad de Ashby, como se ilustra en la figura 2.

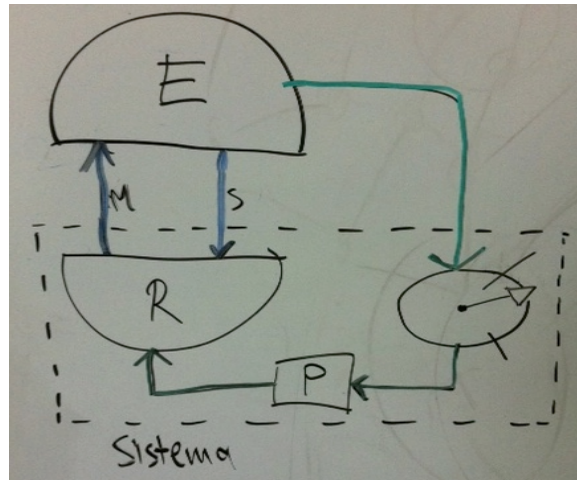


FIGURA 2. Esquema de un sistema caracterizado por la ultraestabilidad.

La parte del sistema responsable de regular la interacción sensoriomotora es representada por R, mientras el ambiente externo es representado por E. Las dos partes están acopladas mediante movimientos (M), desempeñados por el sistema, y las sensaciones (S) recibidas por su entorno. Esto puede ser la base de un bucle de retroalimentación negativa. La forma precisa de comportamiento está determinada por parámetros P, indicados por una flecha que va de P a R. El sistema también incluye una variable esencial, que interpretamos como la temperatura de nuestro cuerpo. Este tipo de variable esencial es ilustrada como un dial con una flecha que tiene que quedarse entre dos límites marcados en su orilla.

El gran logro en el entendimiento de Ashby, fue que la ultraestabilidad podía ser implementada introduciendo un segundo bucle de retroalimentación, conectando el estado de la variable esencial con los parámetros P. En particular, cada vez que la variable esencial se escapa de sus límites, el valor de uno de los parámetros P es modificado aleatoriamente. Este diseño de retroalimentación doble tiene implicaciones profundas en términos de estabilidad.

Imagine que el termostato descompuesto en la casa ha sido remplazado por un nuevo sistema diseñado para ser ultraestable. Ahora, ¿Qué pasa si una vez más cambiamos traviesamente el cableado? Eventualmente la variable esencial, la temperatura, va a salirse de su rango de variabilidad. Y esto a su vez detona un cambio aleatorio en uno de los parámetros del sistema. En este punto, dos cosas pueden pasar: O bien este cambio en la organización del comportamiento del sistema lleva a una mejora, y en este caso la variable esencial regresa a sus límites, o la situación continúa siendo insostenible, y en dado caso habrá otro cambio en la organización del sistema. Efectivamente, el sistema continuará cambiando la forma de su comportamiento hasta que una nueva organización estable haya sido encontrada.

De la ultraestabilidad a la autopoiesis

Hemos visto como los cibernéticos han tratado de explicar el comportamiento intencionado, dirigido a metas intrínsecas, y la auto-adaptación interna en términos de diferentes tipos de retroalimentación. Pero aún existe una diferencia crucial entre estos sistemas artificiales y los seres vivos, que está relacionado con el origen de su identidad que los define como el tipo de

sistemas que son. Aunque algunos artefactos pueden auto-regularse y auto-adaptarse, la noción de su “sí mismo” se mantiene elusiva. En el análisis final, aún reciben su identidad y existencia por parte de los ingenieros que los construyeron y diseñaron, y en dónde trazamos su frontera es arbitrario.

Un organismo, por otro lado, está obligado mantener su identidad física mediante la regulación de sus procesos metabólicos y adaptativos, que se remontan a una forma de auto-producción de su propia identidad como un sistema. Fue otra gran contribución Latino-Americana el resaltar esta importante diferencia entre artefactos cibernéticos y sistemas vivos, y el de darle una nueva etiqueta, “autopoiesis” [3]. Para Maturana (figura 3) y para Varela, que era el estudiante de Maturana en la época, la autopoiesis era la fundación sistémica de la autonomía biológica. Auto-producción significa que la identidad y el comportamiento de un ser vivo están intrínsecamente determinados por ese ser vivo en sí mismo. En el caso de un ser vivo, pero no de los mecanismos artificiales, ser y hacer son interdependientes [6].

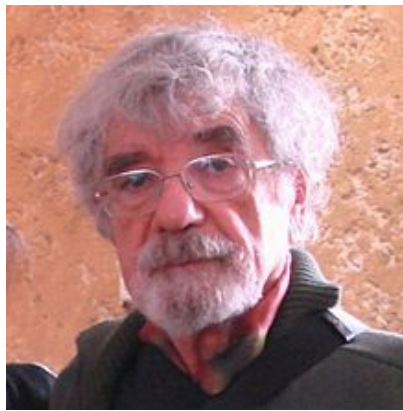


FIGURA 3. Dr. Humberto Maturana
(Foto tomado de http://en.wikipedia.org/wiki/Humberto_Maturana)

Maturana y Varela tenían la intención de que su teoría de la autopoiesis fuera solamente sobre la organización de los seres vivos, sin comprometerse a cualquier especulación sobre las potenciales consecuencias subjetivas de esta organización. Sin embargo, durante la era de la cibernética había también un movimiento en la filosofía continental, que sostenía que la perspectiva del sujeto (su fenomenología), podía estar establecida en la constitución corpórea de lo vivo. Desde la perspectiva de la nueva ciencia cognitiva, el filósofo alemán Hans Jonas ha provisto de algunas de las más importantes contribuciones al respecto [7]. Argumentó que experimentamos un mundo con significado y nuestra existencia nos concierne, porque la corporización biológica es *precaria*. Esencialmente, es porque somos individuos finitos que están constantemente amenazados por la posibilidad de no-existencia, que esta existencia importa en primer lugar. El sentido depende de la mortalidad.

Una de las contribuciones más importantes de Varela (figura 4) hacia el final de su vida fue integrar estas comprensiones fenomenológicas con la teoría sistémica de la autopoiesis, argumentando que de esta forma la ciencia cognitiva será capaz de dar cuenta de la existencia

humana en su totalidad, incluyendo los aspectos tanto objetivos como subjetivos. [8]. Esta propuesta se ha vuelto la fundamentación teórica de la nueva ciencia cognitiva “enactiva”.



FIGURA 4. Dr. Francisco Varela
(Foto tomado de http://en.wikipedia.org/wiki/Francisco_Varela)

Otro desarrollo reciente ha sido regresar a las ideas de la era de la cibernética en cuanto a la auto-regulación y la auto-adaptación. La autopoiesis como concepto categorial no nos dice nada sobre estados que son posibles mientras un sistema se mantiene existiendo tras su auto-producción. Por ejemplo, podemos estar más o menos sanos, más o menos adaptados a las circunstancias. Para entender como esta variabilidad puede ser así, necesitamos combinar la autopoiesis con otro principio, uno que le permita a un sistema medir grados de posibilidades. El científico argentino Di Paolo ha argumentado que esto puede ser logrado haciendo claro que los seres vivos son agentes que tienden a regular sus interacciones sensoriomotoras con relación a sus variables esenciales [4]. Esta idea de la agencialidad está ilustrada en forma de diagrama en la figura 5.

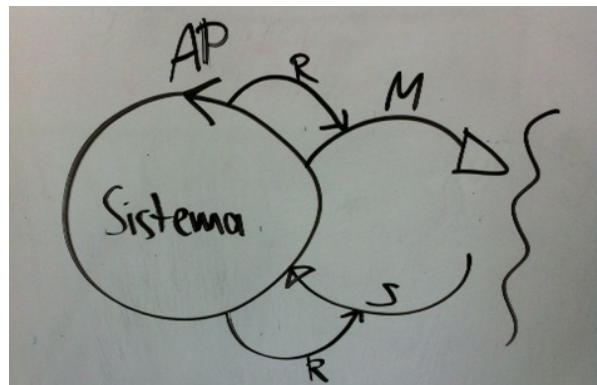


FIGURA 5. Esquema de un sistema caracterizado por agencialidad.

El sistema autopoietico es dibujado como una flecha que se cierra en sí misma, formando así un círculo cerrado. Esto representa la auto-producción (AP). Al igual que en la figura 2 de la ultraestabilidad, el sistema puede afectar el ambiente (línea ondulada) mediante el movimiento (M) y a cambio ser afectado por el ambiente mediante sensaciones (S). Y una vez más el sistema puede regular (R) la forma particular de su acoplamiento sensoriomotor.

Hay por lo tanto muchas semejanzas entre la teoría de ultraestabilidad de Ashby y la teoría de la autopoiesis de Maturana y Varela, pero también hay diferencias importantes [9]. Ashby estaba interesado en entender la capacidad de un sistema de regular sus interacciones, pero esto se mantuvo implícito en el trabajo de Maturana y Varela hasta que Di Paolo de nuevo enfatizó su importancia. Por otra parte, la pregunta sobre la identidad de un sistema se mantuvo implícita en los trabajos de cibernéticos como Wiener y Ashby, hasta que fue hecha prominente por Maturana y Varela. Finalmente, mientras Maturana y Varela inicialmente rechazaron cualquier interpretación subjetiva de la organización biológica, al final Varela se convenció de que la auto-producción y su precariedad son la clave para entender que somos individuos que llevan a cabo vidas significativas.

Conclusiones

Los últimos desarrollos en la ciencia cognitiva se establecen en una larga tradición de teorías revolucionarias que comenzaron en la primera mitad del siglo XX con lo que Wiener llamó cibernética [10]. Como hemos visto, ha habido una contribución sustancial de Latino-América a esta historia, en particular por Rosenblueth, Maturana y Varela, y más recientemente por Di Paolo. Ellos han ayudado a avanzar la vigente búsqueda de una teoría de la subjetividad humana que logre hacer justicia tanto a nuestra existencia vivida (subjetiva) y viviente (objetiva). Al final resulta que la mortalidad no es una consecuencia secundaria de nuestra corporización biológica, que pudiera ser potencialmente evitada (como en el caso de artefactos cibernéticos), pero precisamente aquello que en última instancia nos permite ser individuos interesados por el mundo y a los que les importan otros.

Esto abre un espacio para consideraciones éticas. La nueva ciencia cognitiva toma con seriedad que participamos directamente en las vidas de otros, por ejemplo por medio de la co-regulación de nuestras interacciones sociales [11]. Paso a paso la nueva ciencia cognitiva está poniéndose de acuerdo con toda la complejidad de la existencia humana.

Bibliografía y para saber más

1. Rosenblueth AN, Wiener N, Bigelow J (1943) Behavior, purpose and teleology. *Philosophy of Science* 10: 18-24.
2. Varela FJ, Thompson E, Rosch E (1992) *De Cuerpo Presente: Las Ciencias Cognitivas y la Experiencia Humana*. Barcelona: Gedisa.
3. Maturana HR, Varela FJ (1973) *De Máquinas y Seres Vivos: Una Teoría Sobre la Organización Biológica*. Santiago de Chile: Editorial Universitaria.
4. Di Paolo EA (2015) El enactivismo y la naturalización de la mente. En: Pérez Chico D, Bedia MG, editores. *Nueva Ciencia Cognitiva: Hacia una Teoría Integral de la Mente*. Zaragoza: PUZ.
5. Ashby WR (1960) *Design for a Brain: The Origin of Adaptive Behaviour*. London: Chapman & Hall.
6. Froese T, Ziemke T (2009) Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence* 173: 366-500.
7. Jonas H (2000) *El Principio Vida: Hacia una Biología Filosófica*. Valladolid: Trotta.
8. Froese T (2011) From second-order cybernetics to enactive cognitive science: Varela's turn from epistemology to phenomenology. *Systems Research and Behavioral Science* 28: 631-645.

9. Froese T, Stewart J (2010) Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. *Cybernetics & Human Knowing* 17: 83-106.
10. Froese T (2010) From cybernetics to second-order cybernetics: a comparative analysis of their central ideas. *Constructivist Foundations* 5: 75-85.
11. De Jaegher H (2014) Enacción y autonomía: cómo el mundo social cobra sentido mediante la participación. En: Casado da Rocha A, editor. *Autonomía con Otros: Ensayos sobre Bioética*. Madrid: Plaza y Valdés. pp. 111-131.

Agradecimientos

El autor agradece el apoyo parcial de la DGAPA-UNAM (Dirección General de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México) y del CONACYT (Consejo Nacional de Ciencia y Tecnología) para poder realizar esta reseña, y a Héctor Gómez por su ayuda con la traducción al español.

Ficha bibliográfica

El Dr. Tom Froese es investigador en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la UNAM. Anteriormente, de 2012 hasta 2014 fue becario posdoctoral en el mismo instituto. También cuenta con estancias de posdoctorado en la Universidad de Tokio, Japón, de 2010 a 2012, y en la Universidad de Sussex, Inglaterra, en 2010. El Dr. Froese obtuvo la Maestría de Ingeniería en Ciencia de la Computación y en Cibernética en la Universidad de Reading, Inglaterra, en 2004, y recibió el Doctorado en Ciencia Cognitiva por parte de la Universidad de Sussex, Inglaterra, en 2010. En su tesis de doctorado aplicó el método de la robótica evolutiva para resolver problemas teóricos en las ciencias de la mente. Actualmente su investigación está enfocada en entender mejor los aspectos distintivos de la mente humana relacionados con la cultura y la conciencia.