La equivalencia formal en el lenguaje de las neuronas lógicas de McCulloch y Pitts

Rocío Stefanazzi Kondolf[1]

Recibido: 15 de agosto de 2021
Aceptado:17 de agosto de 2022

Resumen: En este artículo reflexionamos sobre el lugar de la equivalencia formal en el lenguaje propuesto por Warren McCulloch y Walter Pitts en su "A logical calculus of the ideas immanent in nervous activity" ["Un cálculo lógico de las ideas inmanentes en el sistema nervioso"]. Estudiamos el modelo a partir de los aportes que este ha significado en la historia de la ciencia: como teoría computacional de la mente; y como formalismo que contribuyó al desarrollo de la teoría de los autómatas y el diseño lógico. Estudiamos este *paper* desde la teoría de modelos y desde la teoría de la explicación mecanicista en ciencias. A su vez, reflexionamos acerca del carácter de cálculo lógico del texto en cuestión. Consideramos la relación entre el modelo y su *target* y ofrecemos dos interpretaciones posibles: las neuronas lógicas son el modelo que representa a la mente-cerebro o se trata de un nuevo objeto conceptual que posibilita el desarrollo de los autómatas artificiales. Observamos cómo la noción de equivalencia formal juega un papel fundamental: como explicación científica de una teoría de la mente, como criterio de isomorfismo de un modelo de la mente, y como criterio de identidad de las neuronas lógicas o neuronas formales.

Title: Formal equivalence in the language of McCulloch's and Pitts' logical neurons

Palabras clave: equivalencia formal, neuronas lógicas, nuevo necanicismo, modelos científicos.

Abstract: In this article we consider the significance of formal equivalence in the language proposed by Warren McCulloch and Walter Pitts in their original "A logical calculus of the ideas immanent in nervous activity". We study the model from the contributions that it has meant in the history of science: as a computational theory of mind; and as a formalism that contributed to the development of automata theory and logical design. This question is framed into the mechanistic explanation account and the philosophy of scientific models, while also attending to how the model embedded a conception of a language as a logical calculus. We consider the relationship between the model and its target system, offering two possible interpretations: logical neurons are the model that represents the mind-brain or it is a new conceptual object that enables the development of artificial automata. We observe how the formal equivalence notion plays a fundamental role: as scientific explanation about mind's theory, as an isomorphism criterion of mind's model, and as identity criterion of logic neurons or formal neurons.

Keywords: formal equivalence, logical neurons, new mechanism, scientific models.

1. Introducción

En este texto reflexionamos sobre el rol de la equivalencia formal propuesta en el *paper* "A logical calculus of the ideas immanent in nervous activity" de Warren McCulloch y Walter Pitts escrito en 1943.[2] Resulta relevante reflexionar acerca de este *paper* por el lugar

[1] Universidad de Buenos Aires. Buenos Aires, Argentina.
[2] En este trabajo utilizamos la reedición de 1990 en el *Bulletin of Mathematical Biology* (McCulloch & Pitts, 1943/1990). El original se publicó en el *Bulletin of Mathematical Biophysics* (McCulloch & Pitts, 1943).

✉ rociosk3@gmail.com | 0000-0002-2755-0016

Stefannazi Kondolf, R. (2022). La equivalencia formal en el lenguaje de las neuronas lógicas de McCulloch y Pitts. *Epistemología e Historia de la Ciencia, 7*(1), 22–40.
https://revistas.unc.edu.ar/index.php/afjor/article/view/34428/

histórico que ocupa en la constitución de la cibernética (Abraham, 2002; Dupuy ,2000), su aporte a la teoría de la mente y sus influencias en el desarrollo de un lenguaje que contribuyó al surgimiento de las ciencias de la computación, los estudios de inteligencia artificial (IA) (Abraham, 2002; Piccinini, 2003,) y al desarrollo de las teorías del procesamiento de información (Abraham, 2002, pp. 3–4).

Este texto ha sido un puntapié para las investigaciones en la teoría de los autómatas y la arquitectura de von Neumann quien se inspiró en dicho trabajo en su "First Draft of a Report on EDVAC" (1945) y en "The General and Logical Theory of Automata" (1951). Algunos consideran que también sirvió de "modelo base" para la creación del *perceptron* de Frank Rosenblatt (Dupuy, 2000; Piccinini, 2003; Zamora-Cárdenas, Zumbado, Trejos-Zelaya, 2020, p. 17). En cuanto a la teoría de la mente, el modelo ha tenido influencia en el desarrollo de las ciencias cognitivas y las neurociencias (Abraham 2000).

El lenguaje que McCulloch y Pitts crearon ha sido leído por algunos autores, como es el caso de Gualtiero Piccinini (2004), como un mecanismo para explicar procesos de conocimiento y no el sistema nervioso. Dicho autor recorre las búsquedas teóricas y las experiencias de McCulloch y Pitts a los fines de comprender qué significó el trabajo de 1943 en tanto episodio de la historia de la ciencia. En este sentido, nos indica cuatro aportes: (1) un formalismo que permite la generalización de la noción de autómata finito, (2) una técnica inspirada en la noción de diseño lógico, (3) el primer uso de una computación de manera conjunta al problema mente-cuerpo, y (4) la primera teoría computacional de la mente y el cerebro. Que Piccinini (2004) considere que la propuesta del texto en cuestión sea un *mecanismo* para explicar el funcionamiento procesos mentales como el conocimiento u otros, está sostenido en la trayectoria de investigación de McCulloch, quien había trabajado en la teoría de los psicones. Los psicones eran algo así como "átomos mentales" equivalentes a proposiciones sobre sus antecedentes temporales. La teoría de los psicones fue clave para la resolución propuesta del problema mente-cuerpo: a cada átomo mental correspondía una actividad neuronal. A la equivalencia psicón-actividad neuronal se agrega la equivalencia con la actividad proposicional. El mecanismo de la mente es el mecanismo del cerebro y estos a su vez coinciden con el mecanismo de una lógica proposicional bivalente. A su vez, si el mecanismo puede ser considerado un mecanismo computacional, entonces para Piccinini (2003, 2004) la teoría de McCulloch y Pitts es una teoría computacional de la mente.

Por otro lado, autores como Mark Schlatter y Ken Aizawa (2008) ponen el foco en los aspectos formales-matemáticos, imbricados a la historia y las búsquedas teóricas de los investigadores, pero con más énfasis en la trayectoria de Pitts. En este sentido, consideran que McCulloch aportó la idea del uso del álgebra de Boole inspirado en el mecanismo "todo-nada" del modelo de neuronas biológicas. A esto Pitts lo conjugó con un lenguaje de funciones, que era un producto de sus trabajos en las redes neuronales de Alston Householder; y una escritura lógica de segundo orden, dada por el lenguaje II de Rudolf Carnap, con profunda influencia de Bertrand Russell. A esto que indican Schlatter y Aizawa, podríamos sumarle, además, una organización del texto mismo en términos de un cálculo o lenguaje lógico, probablemente también por la influencia de Carnap y, previamente, de Russell.

En este artículo retomamos los cuatro aportes del *paper* de McCulloch y Pitts que resume Piccinini (2004) mencionados anteriormente y, a partir de allí, evaluamos los sentidos de equivalencia formal, y el rol que juega esta noción en la articulación de cada

uno de dichos aportes. Utilizamos el marco de la explicación mecanicista para comprender el sentido de "mecanismo" que Piccinini pone de manifiesto en las teorizaciones de la mente de McCulloch. A su vez, utilizamos la teoría de modelos para pensar la relación representativa del modelo McCulloch-Pitts en cuanto al mecanismo de la mente y el diseño lógico de computadoras. Consideramos que si bien el modelo fue pensado para representar la mente-cerebro, en sus intentos de abstraer lo que concebían como su mecanismo lógico inmanente, posibilitaron la construcción de nuevas entidades conceptuales: las neuronas lógicas o formales. En este punto surge una cuestión doble, o las neuronas lógicas son el modelo de la mente-cerebro que sería el *target* u objetivo de aquel, o se trata entonces de un nuevo fenómeno *target* que posibilitó el desarrollo de la teoría de los autómatas artificiales.

Nos centramos en la primera parte del artículo de 1943 relativo a las redes neuronales sin círculos y mencionamos solo algunas cuestiones de la segunda parte, relativas a las redes con círculos. Reflexionamos sobre el lugar de la equivalencia formal en la constitución de las neuronas lógicas o neuronas formales del lenguaje propuesto por Warren McCulloch y Walter Pitts en su artículo. Observamos cómo la noción de equivalencia formal juega un papel fundamental: como explicación científica de una teoría de la mente, como criterio de isomorfismo de un modelo de la mente, como criterio de identidad de las neuronas lógicas. Insertamos esta cuestión en el marco de las explicaciones mecanicistas en filosofía de las ciencias, la reflexión contemporánea sobre modelos científicos y la concepción de los lenguajes como cálculos lógicos.

2. Los recorridos e influencias lógico-matemáticas

En este apartado desarrollamos los recorridos e influencias en el ámbito matemático, lógico y computacional Warren McCulloch y Walter Pitts mientras que en apartados posteriores nos dedicaremos a reconstruir las influencias de la neurofisiología.

Como nos indica Piccinini (2004), la formación y dedicación principal de McCulloch era la de psiquiatra y neurofisiólogo. Sus primeras investigaciones postulaban unas entidades que ya mencionamos: los "psicones"[3], una especie de átomos mentales que, luego asociados a expresiones proposicionales, le permitieron pensar una teoría de la mente y del conocimiento. Es en este contexto, según nos indica Piccinini, que McCulloch piensa la aplicación del álgebra booleana a las redes nerviosas.

En su búsqueda de una correlación lógica para sus átomos mentales, McCulloch se acercó al seminario filosófico en el cual Frederic Fitch dictó unas clases sobre los *Principia Mathematica* de Russell y Whitehead, en la Universidad de Yale. En este mismo marco, se acercó al trabajo de Nicolas Rashevsky (con quien trabajaba Pitts) a través Filmer Northrop —director del seminario filosófico— quien reivindicaba la importación de técnicas y formalismos de la física-matemática a la biología; en 1937, también en Yale, se acercó al trabajo de Joseph Henry Woodger acerca del uso de los sistemas axiomáticos en biología.[4] A

[3] El término en inglés es "psychons".
[4] Abraham (2002) rastrea en Northrop (1931): "La etapa descriptiva y clasificatoria en biología que comenzó con Aristóteles había comenzado a moverse hacia una etapa formal y deductiva con el trabajo de Joseph Henry Woodger (1894–1981) (Woodger, 1937) y Nicolas Rashevsky (1899–1972). En estos trabajos, la lógica formal y las matemáticas jugaron un papel importante, por lo que la biología como disciplina fue alcanzando una etapa más madura, en tanto que comenzó a incorporar el método científico de la física, es decir, utilizando el análisis teórico y las formulaciones matemáticas. Los argumentos de Northrop sobre el valor de la formalización en biología estaban conectados con su visión más amplia de 'diseccionar las teorías científicas dadas que. . . los científicos han verificado, para determinar qué conceptos y principios se toman como primarios o indefinidos' (Northrop, 1931, p. xiii)" (Abraham, 2002, pp. 6–7).

principios de 1940 conoce los trabajos de Alan Turing sobre las máquinas universales[5]. Un año después conocerá el trabajo de Rashevsky en biología matemática en la Universidad de Chicago y, este contexto, conocerá el trabajo de Pitts.

Schlatter y Aizawa (2008) nos indican que Pitts comenzó sus investigaciones en el grupo de Rashevsky en la Universidad de Chicago y, en particular, a partir de los trabajos de Householder (1941, 1942). La diferencia importante del *paper* de 1943 con el enfoque de Householder (1941a, 1941b, 1941c, 1942), se debió a que, en vez de trabajar con matemáticas continuas, lo hizo con matemáticas modulares o discretas.[6] Sin embargo, Pitts ya había introducido varios cambios a las explicaciones de Householder en una serie de artículos que había escrito previamente. Mientras este último había planteado la relación entre las neuronas en términos de sistemas de ecuaciones, Pitts planteó la estimulación en función del tiempo. De esta manera, resolvió el problema planteado por su antecesor: dada una red con un patrón de estimulación específico, encontrar el patrón de excitación. Otro cambio que produjo fue pensar en una sinapsis a lo largo del tiempo en vez de en la red en su totalidad. Exploró cómo cambia la estimulación de las sinapsis y, con el paso del tiempo, se acerca al estado estacionario [*steady-state*[7]]. A su vez, consideró otras posibles redes topológicas y distinguió entre sinapsis de 1er (una fibra de entrada y una de salida) y de 3er orden (una fibra de entrada y más de una de salida). Por último, para dar cuenta de los valores posteriores de estimulación, introdujo el operador ***E*** que se comporta de la siguiente manera:

- $\mathbf{E\,f(x) = f(x + 1)}$**:** El operador ***E*** remite a la estimulación una unidad de tiempo después de x
- $\mathbf{E^3\,f(x) = f\,(x + 3)}$**:** El operador ***E*** elevado a la tres remite a la estimulación tres unidades de tiempo después de x

Este operador es importante porque luego se utilizará uno semejante en el *paper* de 1943, llamado operador **S**.

Una segunda influencia muy importante para Pitts fue la de Rudolf Carnap a quien conoció por su interés en el trabajo de Russell y Whitehead. En su adolescencia había leído los *Principia Mathematica*, luego se acercó a unas conferencias que Russell dictó en Chicago, y éste lo contactó con Carnap. De Russell y de Carnap tomarán la idea de cálculo lógico, y en particular de Carnap una escritura lógica de segunda orden, dada por su lenguaje II.

---

[5] Hoy conocidas como máquinas de Turing. Tal como la define Turing (1936) una máquina de computar [*computing machine*] es aquella que tiene una serie finita de condiciones $q_1, q_2, \ldots, q_n$ las cuales se llaman m-configuraciones [*m-configurations*] y una cinta dividida en secciones [*squares*] las cuales pueden llevar un símbolo (0 ó 1). Una de esas secciones es el escáner de la máquina [*scanned square*] y el símbolo en dicha sección se llama símbolo escaneado [*scanned symbol*]. El símbolo escaneado es el único del cual la máquina es "directamente consciente", aunque si se cambian algunas configuraciones, ésta puede recordar algunos símbolos que ya han sido escaneados previamente. El comportamiento de la máquina queda determinado en cualquier momento por las configuraciones y el símbolo escaneado. Si la sección está en blanco la máquina puede escribir un símbolo y si tiene un símbolo puede borrarlo. La máquina solo cambia desde la posición que ha sido escaneada y siempre cambia de a un lugar moviéndose hacia la derecha o a la izquierda. Una máquina universal es aquella que puede computar cualquier secuencia computable que computen otras máquinas posibles (Turing, 1936, pp. 231, 232, 241, 242).

[6] Es decir, aquella que trabaja con funciones con cortes, que no son continuas, como es el caso de la lógica y el álgebra.

[7] "La actividad en estado estacionario [*steady-state*] bajo estimulación constante es la más simple de considerar [...]. Es decir, si se aplica un estímulo constante durante un período de tiempo a cada una de las fibras del complejo, lo que sea que las fibras hagan en los primeros milisegundos, o incluso segundos, suponemos que establece un breve estado de respuesta que persiste sin cambio mientras los estímulos se mantienen sin cambios" (Householder, 1941a, p. 64).

En síntesis, las influencias han sido varias: Por un lado, Northtrop, Woodger, Rashevsky y Householder en cuanto al uso de la matemática en biología; Boole, Russell y Carnap en cuanto a las influencias lógico-matemáticas y Turing en cuanto a la concepción de computabilidad.

3. Una teoría de la mente

En este apartado seguimos la reconstrucción que hace Piccinini (2003, 2004) de la teoría de la mente de McCulloch y Pitts teniendo en cuenta las siguientes cuestiones que fueron mencionadas anteriormente: que se trata de la primera teoría computacional de la mente; que es una teoría mecanicista de la mente; que necesitó de una concepción de entidades atómicas llamadas psicones —desarrollada por los trabajos anteriores de McCulloch— para establecer una equivalencia con la actividad neuronal; y que, de este modo, dio una respuesta al problema mente-cuerpo con la identificación de la mente y el cerebro.

3.1. La primera teoría computacional de la mente

Para Piccinini (2004) la primera teoría de la mente de McCulloch y Pitts es la primera teoría computacional en tanto utiliza la noción matemática de computación de Turing, incluso antes que éste último esbozara opiniones sobre la relación de su teoría computacional con la teoría de la mente (p. 176). Sin embargo, que hayan usado la noción de computación de Turing, no significa que las redes neuronales podrían computar cualquier cosa que compute una máquina de Turing, lo cual ha sido un error usual de interpretación (Piccinini, 2004, p. 177). Este error puede rastrearse en la influencia que McCulloch y Pitts tuvieron sobre la interpretación de la tesis Church-Turing. En la apelación a Turing que hacen en su texto aparece la siguiente idea:

> Se muestra fácilmente: primero, que cada red, si está equipada con una cinta, escáneres conectados a aferentes y eferentes adecuados para realizar las operaciones motoras necesarias, puede calcular sólo tales números como una máquina de Turing; segundo, que cada uno de los últimos números puede ser computado por tal red; y que las redes con círculos pueden ser computadas por tal red; y que las redes con círculos pueden computar, sin escáneres y una cinta, algunos de los números que la máquina puede, pero no otros, y no todos. Esto es de interés como una justificación psicológica de la definición de Turing de computabilidad y sus equivalentes, la definibilidad λ de Church y la recursividad primitiva de Kleene:[8] si cualquier número puede ser calculado por un organismo, es computable por definición, y de manera conversa. (McCulloch & Pitts, 1943/1990, p. 113)

Es decir, no están diciendo que sus redes sean matemáticamente equivalentes a la noción de computabilidad de Turing, como usualmente se ha comprendido erróneamente, sino que estas redes son una justificación psicológica de la posibilidad lógica de la tesis de Turing-Church. Tal como señala Piccinini (2003):

> La conexión trazada por McCulloch y Pitts, a través de sus redes, entre los fenómenos mentales y la computación ha tenido efecto en el modo que se interpretó la Tesis de Church-Turing (TC). Según la teoría de McCulloch y Pitts, cada red podría ser descripta como computando una función. Entonces, en el sentido en que las redes de McCulloch-

[8] Posteriormente Kleene (1956) retomó las cuestiones planteadas por McCulloch y Pitts y mostró que este formalismo era equivalente a los autómatas finitos.

> Pitts computaban, y en la medida en que las redes de McCulloch-Pitts eran un buen modelo del cerebro, cada actividad neuronal era una computación. Esto fue particularmente significativo dado que McCulloch y Pitts consideraban que las computaciones de sus redes eran explicaciones de los procesos mentales. (Piccinini, 2003, p. 63)

La teoría de McCulloch-Pitts parecía ser entonces una fundamentación psicológica de los límites de lo computable pero no así una tesis sobre los límites lógicos de lo computable.

3.2 Una teoría mecanicista de la mente

El llamado nuevo mecanicismo es una corriente heterogénea de pensadores y posturas en torno a la definición de mecanismo, pero comparten la reflexión común acerca de dicha noción. Las explicaciones mecanicistas surgen en el contexto de las llamadas *ciencias frágiles* (Barberis, en prensa) y no pretenden una explicación general de una teoría sino explicaciones locales y situadas de ciertos fenómenos en términos de mecanismos. Este tipo de explicaciones, como nos indican Branca, Ramírez, y Vilatta (2015), discuten con el reduccionismo que ha presentado a las explicaciones científicas como forma de reducir una teoría a otra. Craver (2007) ha indicado que la tradición reduccionista construye las explicaciones a través de enunciados de identidad en los que un nivel se deriva de otro y se requiere un mapeo uno-a-uno (pp. 107–108). En este sentido, la concepción mecanicista de McCulloch-Pitts establece enunciados de identidad que no son reduccionistas sino que todos los niveles involucrados en el mapeo uno-a-uno tienen la misma jerarquía.

En este punto seguimos el rastreo del mecanicismo que Piccinini (2004) hace de la teoría de la mente de McCulloch y Pitts a partir de las concepciones de los trabajos previos de McCulloch, para quien uno de los objetivos y aportes de la neurofisiología ha sido explicar la mente "en términos de mecanismos neuronales" (p. 177). En esta dirección, Dupuy (2000) nos dice que "McCulloch […] fue un hombre poseído por una sola idea, la cual dependía de sostener al mismo tiempo cerebro, mente y máquina" (p. 111).

Tal como señalamos anteriormente, Piccinini (2004) encuentra en el propio discurso del neurofisiólogo una teoría de la mente que suponía átomos mentales —los llamados psicones—, las entidades mentales equivalentes a la actividad proposicional de las neuronas:

> […] un psicón es ''equivalente'' a una proposición sobre su antecedente temporal. En terminología más reciente, McCulloch parecía pensar que un psicón tiene un contenido proposicional, que contiene información sobre la causa de ese psicón. Un segundo punto clave fue que un psicón "propone" algo a un psicón posterior. Esto parece significar que el contenido de los psicones se puede transmitir de un psicón a otro, generando "los equivalentes" de proposiciones más complejas. (Piccinini, 2004, p. 178)

McCulloch comenzó a elaborar su teoría de los psicones a mediados de 1920. En 1928 hizo una estancia en Nueva York en el hospital Bellevue en el cual trabajó en la teoría de la función nerviosa y a partir de la cual pensó que ciertos problemas como la epilepsia podrían deberse a bucles o *loops* en la actividad nerviosa, y creía encontrar sustentos en las investigaciones de Ramón y Cajal. Una influencia importante en sus redes con *loops* fue el neurólogo y psicoanalista Lawrence Kubie (Dupuy, 2000, p. 55). Estos bucles o *loops*, asociados a problemas como la epilepsia o la enfermedad de Parkinson, reaparecerán en el artículo de 1943. En 1929 comienza a pensar en la hipótesis de que los impulsos eléctricos todo-nada de las neuronas podrían corresponderse a los psicones y que "las relaciones de excitación e inhibición entre las neuronas realizarían operaciones lógicas sobre señales

eléctricas correspondientes a inferencias de su cálculo proposicional de psicones." (Piccinini, 2004, p. 179). Luego, en 1934, trabajó con Joannes Dusser de Barenne en su laboratorio de Neurofisiología en el mapeo de las conexiones entre áreas del cerebro. (Piccinini, 2004, pp. 178, 179, 180). Es en el discurso de la neurofisiología, en la equivalencia entre neuronas biológicas y psicones, que McCulloch encuentra la justificación de su concepción mecanicista de la mente y a partir de las preguntas que surgen en el seno de la neurofisiología incorpora nuevas disciplinas, en particular, la lógica, la matemática, la psicología.

En resumen, la estrategia de McCulloch y Pitts, según la lectura de Piccinini (2004) que venimos siguiendo, fue: (a) en primer lugar, reducir y simplificar las redes neuronales, y, de esta manera, una serie de inferencias proposicionales podían ser mapeadas en eventos neuronales y viceversa. En segundo lugar (b), asumir que los pulsos neuronales tenían contenido proposicional que se correspondían con procesos mentales atómicos. Esta teoría de la mente asociada a las expresiones proposicionales a través de la dinámica misma del impulso nervioso necesitó del supuesto mente = cerebro disolviendo de algún modo, o dando una respuesta al problema mente-cuerpo.

4. Un modelo de la mente-cerebro

Las teorizaciones de McCulloch y Pitts han producido un modelo que, a la hora de construir su *target*, tomó como punto de partida una serie de asunciones o supuestos "físicos", tal como los nombran los propios autores del modelo. Potochnik, Colombo y Wright (2019) definen que "Una asunción […] es una especificación que un sistema *target* debe satisfacer para un modelo dado, de ser similar en el modo esperado", y que "Las asunciones pueden ser idealizaciones que no necesariamente serán verdaderas." (p. 99). Es decir, que un *target* no necesariamente se corresponderá con el fenómeno real. Aunque los autores mencionados anteriormente tienden a considerar al *target* como el fenómeno real (Potochnik, Colombo, & Wright, 2019, p. 96). En contraposición a este tipo de posturas realistas, Cassini (2018) considera que el *target* de un modelo no es el fenómeno real, sino el producto de un proceso de construcción científica. Lo que nos interesa principalmente en este apartado es rastrear las asunciones que se han abstraído e idealizado[9] en el modelo que nos posibilitarían rastrear el *target* del mismo.

4.1. La neurofisiología teórica y los "supuestos físicos"

En el artículo de 1943 McCulloch y Pitts toman el modelo de la neurona de lo que llaman la "neurofisiología teórica", sobre la cual no citan ninguna bibliografía.[10] En la "Introducción" describen al sistema nervioso como una red de neuronas a las cuales le atribuyen dos grandes partes: el soma y el axón. Es notable que no toman en cuenta a las dendritas. El axón de una neurona se conecta con el soma de otra mediante la sinapsis, a través de impulsos nerviosos que pueden ser inhibitorios o excitatorios. Hay un período de adición latente en el que la neurona puede recibir impulsos. Para que se inicie un impulso nervioso, la excitación de la neurona debe exceder el límite o umbral que está determinado por la neurona misma. Una vez iniciado el impulso, se propaga desde el punto de excitación hacia

[9] "Omitir o ignorar ciertas características conocidas del sistema es una abstracción; incluir características que el sistema target no tiene es una idealización" (Potochnik, Colombo, & Wright, 2019, p. 118)
[10] Michael Arbib indica que parten de la doctrina de la neurona de Ramón y Cajal y la sinapsis de Sherrignton. (1987, p. 4)

el resto de la neurona. Entre la recepción de un impulso y la propagación de otro hay un retraso sináptico (McCulloch & Pitts, 1943/1990). Respecto a la excitación es interesante señalar dos cosas. En primer lugar, que los autores plantean un debate en torno a su causa, pero no la consideran relevante para el cálculo. Esto nos da la pauta de que sólo toman algunas características del modelo de la neurofisiología para crear el formalismo. En segundo lugar, que la característica que aporta la excitación es la direccionalidad del impulso nervioso que se mueve siempre en una misma dirección espacial y temporal.[11]

A continuación presentamos un esquema que ilustra las partes y relaciones de las redes neuronales "biológicas" que los autores retoman:

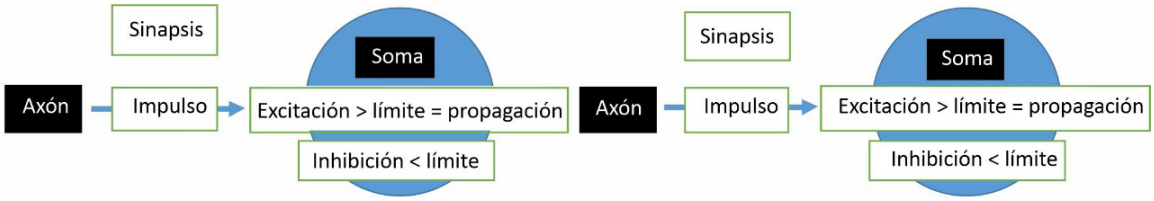

Figura 1.

Esta presentación de algunos conceptos neurofisiológicos, es retomada en la segunda parte del texto relativa a las redes sin círculos. Allí hacen una presentación, quizás a modo de sistematización, de los "supuestos físicos" para el cálculo:[12]

1. La actividad de la neurona es un proceso todo-o-nada.
2. Un cierto número fijo de sinapsis deben ser excitadas dentro de un período de adición latente para excitar a una neurona en cualquier momento, y su número es independiente de la actividad previa y la posición de la neurona.
3. El único retraso significativo en el sistema nervioso es el retraso sináptico.
4. La actividad de una sinapsis inhibitoria previene absolutamente la excitación de una neurona en ese instante.
5. La estructura de la red no cambia con el tiempo. (McCulloch & Pitts, 1943/1990, p. 101)

En lo que respecta al cálculo, el punto (1) será el puntapié de la inspiración. Que las neuronas se comporten de dicha manera posibilita la equivalencia con una lógica bivalente (con 0 y 1 como valores de verdad) en la que se toma como válido el Principio del Tercero Excluido. En relación al punto (2), lo importante es que nos indica que hay una "sumatoria" de las sinapsis previas cuyo resultado será el estímulo que recibirá el soma de otra. Si pasa su límite —independiente del estímulo[13]— el impulso es excitatorio. El punto (3) refiere a la dimensión temporal de las sinapsis. El ítem (4) nos habla del caso en el que el impulso no excede el umbral. Finalmente, (5) refiere a propiedades estructurales de la red. Según este

[11] Esto en el formalismo será expresado por el teorema 6.
[12] Según Schlatter y Aizawa (2008), la primera sección funciona como una introducción más general a la neurofisiología, lo que se sostiene en la trayectoria e investigación de McCulloch (el uso de la lógica booleana para pensar la relación entre neurona y las investigaciones de neurotransmisión en macacos, junto a Barenne, lo que les llevó a "descubrir" el fenómeno de extinción). La segunda sección lista las asunciones preliminares neurofisiológicas. La única diferencia que encuentra entre la segunda y la primera está en las velocidades de la propagación del potencial de acción. (2008, pp. 242-243)
[13] El rol excitatorio de la neurona postsináptica es independiente de la intensidad del estímulo total ya que, en cierto sentido, "rectifica" los estímulos de entrada.

ítem pareciera que no se admiten cambios en las estructuras de las redes luego de las sinapsis que se han producido.

Este momento del artículo que acabamos de exponer relativo a los supuestos físicos construye el *target* del modelo a partir de la idealización y abstracción de algunos conceptos del discurso de la neurofisiología sobre el mecanismo de comunicación neuronal. En este sentido, y también bajo el supuesto sobre el que hablamos anteriormente acerca de la identificación entre neuronas biológicas y psicones, podemos decir que el *target* del modelo es la mente-cerebro.

5. El formalismo de McCulloch y Pitts

En este apartado damos cuenta de estos objetos conceptuales y su proceso de comunicación en el marco de una descripción general del lenguaje lógico de McCulloch y Pitts. El desarrollo de este lenguaje, nos permitirá, en apartados posteriores, reflexionar acerca de la relación del discurso de la neurofisiología teórica con el de la lógica-matemática y sus efectos en la construcción del *target* del modelo.

Nos centraremos en la presentación del lenguaje que aparece en el marco del estudio de las redes sin círculos [*nets without circles*][14] y presentaremos algunos teoremas. Las redes con círculos tendrían la complejidad de remitirse a tiempos remotamente pasados y requerirían de una cuantificación temporal (Piccinini, 2004, p. 197), mientras que las redes sin círculos se limitan a expresar el tiempo a través de los predicados y los funtores.

Luego de los supuestos físicos, los autores desarrollan un formalismo. La escritura de McCulloch y Pitts es un poco compleja y, según nos indica Schlatter y Aizawa (2008), tiene errores. Haremos una interpretación de su exposición para comprender algunas cuestiones de su lenguaje. A continuación listamos sus componentes:

- ***S*** Funtor[15] definido por $\mathbf{S(P)}\,(t)\,.\equiv.\,(\mathbf{P}\,(\mathbf{Kx}) \wedge t = x')$.[16] "Cuyo valor para una propiedad *P* es la propiedad que vale para un número cuando *P* vale para sus predecesores." (McCulloch & Pitts, 1943/1990, p. 102)
- **Pr** es la expresión-predicado que se usa para hablar de P sin la aclaración de los paréntesis.

Como pasaba con el operador **E** en los trabajos tempranos de Pitts, si elevamos el Funtor *S* a un número, ese número nos indica la cantidad de veces que se aplicó el Funtor *S*: $S^2PR = S\big(S(PR)\big)$ donde *PR* remite a una *expresión-predicado*.

- ***N*** nombre de la red ($\mathfrak{N}$ en el original).
- $c_1, c_2, c_3, \ldots, c_n$ nombre de las neuronas
- N_i denota la propiedad de un número que una neurona c_i dispara a un tiempo que es ese número de retrasos sinápticos desde el origen del tiempo. (N_i^8 es la acción de c_i 8 instantes de tiempo desde el comienzo).

[14] Una idea semejante había aparecido en Turing (1936) para hablar de las máquinas, los términos que aparecen allí son "máquina circular" [*circular machine*] y "máquina libre de círculos" [*circle-free machine*] (p. 233).
[15] Este término lo toman de Carnap (1937/2007): "A los fines de expresar propiedades o relaciones de posición por medio de números, usaremos funtores" (p. 14). Actualmente, la noción de funtor es usada en teoría de categorías para hablar de morfismos entre categorías (ver Lawvere & Schanuel, 1997).
[16] La escritura original es: $\mathbf{S(P)}\,(t)\,.\equiv.\,\mathbf{P}\,(\mathbf{Kx})\,.\,t = x')$. Los puntos en los costados del símbolo de la equivalencia son la manera en la que se representaba la equivalencia. Actualmente se expresa sin los puntos (≡). A su vez, hay un error en el uso del paréntesis.

- $N_i(t)$ es la función que indica que c_i se dispara en un tiempo t
- N la clase de todos los $N_1, N_2, \ldots$
- $\mathbf{N_1}, \ldots, \mathbf{N_P}$ *y* $\mathbf{N_{p+1}}, \ldots, \mathbf{N_n}$, para diferenciar la acción de los aferentes periféricos (aquellas N_p que no tiene axones de entrada) de las acciones del resto de las neuronas (N_{p+1})

Dado el lenguaje anterior, la solución de la red ***N,*** los retrasos sinápticos de la misma, estará dada por oraciones S de la forma:

$$S_i\colon\ N_{p+1}(z_1). \equiv . Pr_i(N_1, N_2\ ,\ .\ .\ .\ ,\ N_P,\ z_1)$$[17]

Lo cual refiere al problema planteado por los autores como "Encontrar un método efectivo para obtener un conjunto **S** calculable o computable que constituya una solución para cualquier red dada". Y de modo converso, dado un **Pr**, se puede encontrar su realizabilidad en la red. Esto refiere al segundo problema: "Caracterizar la clase realizable **S** de una manera efectiva" (McCulloch & Pitts, 1943/1990, p. 103). Aquí puede observarse la inspiración que Turing había producido en los autores del artículo de 1943, ya que la noción de función efectivamente computable[18] surge de los trabajos de aquél.

Luego de plantear estos problemas, nos dan la definición de expresión proposicional temporal (TPE). Esta noción se vuelve importante porque es la que nos indica el tipo de proposiciones legítimas dentro del cálculo, aquellas que nos dictan el modo en que las "neuronas lógicas" se comunican. Una TPE es definida por la designación de una función proposicional temporal (TPF) de manera recursiva:

1. Una $^1p^1\ [z_1]$ es una TPE, donde p_1 es una variable de predicados.[19]
2. si S_1 y S_2 son TPE, entonces también lo son SS_1, $S_1 \vee S_2$, $S_1 \wedge S_2$ *y* $S_i \wedge \sim S_2$
3. Nada más es una TPE. (McCulloch & Pitts, 1943/1990, p. 103)

Como se observa en el ítem 2., estas TPE se forman a partir de las conectivas lógicas conjunción, disyunción y negación.

Luego de esta definición, los autores presentan una serie de teoremas relativos a estas TPE para poder extender su comportamiento de una manera sistémica. Por ejemplo, en el teorema 2 demuestran que las TPE son realizables en una red de orden cero. Primero se prueba para algunas TPE y luego, mediante inducción matemática, se concluye que todas las TPE son realizables.

Este lenguaje tendrá un correlato gráfico semejante a las compuertas lógicas AND, OR y AND-NOT, en cuanto modelo de implementación de funciones proposicionales binarias del álgebra booleana:

[17] Aquí S refiere a la solución de la red, no al funtor definido anteriormente, mientras que z_1 es una variable libre.
[18] Las funciones Turing-computables (es decir, computables por una máquina de Turing son funciones efectivamente computables. Lo que no es obvio es la conversa y que constituye la famosa Tesis de Turing: si una función es efectivamente computable, entonces es Turing-computable. Boolos, Burgess y Jeffrey definen a una función f efectivamente computable como aquella que puede dar una lista de instrucciones para poder determinar el valor de $f(n)$ para cada argumento de n (ver "Turing Computability" en Boolos, Burgess, & Jeffrey, 2007, pp. 3-33).
[19] z_1 según definieron unas páginas antes en el texto es una variable libre y $^1p^1$ es un predicado dentro del conjunto mayor *PR*.

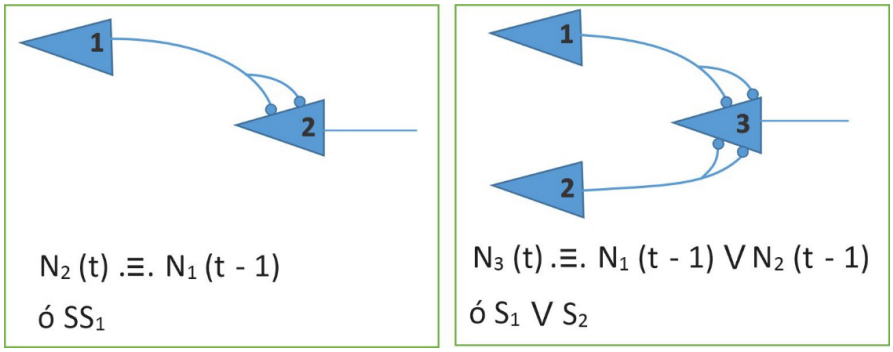

Figura 2.

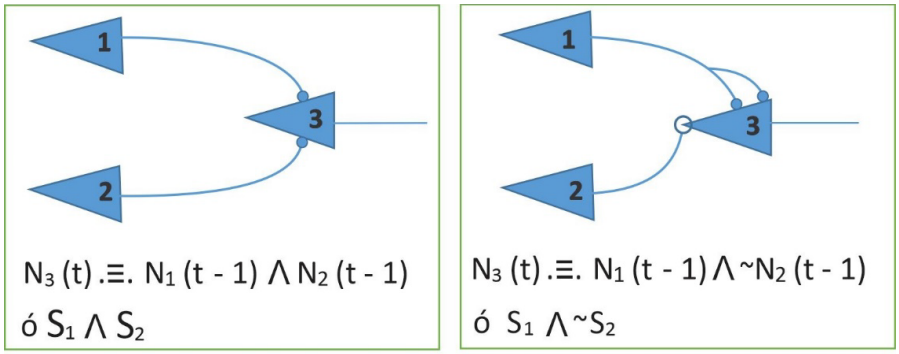

Figura 3.

La escritura en la que se usa la función $N_i(t)$ nos permite indicar el momento en el que se dispara la neurona en relación a un tiempo anterior a t. De modo que esta escritura da cuenta de sus posibles procesos anteriores como equivalentes a la función del disparo en un tiempo t. Mientras que el funtor S nos permite resumir esa expresión temporal. Este nos indica el disparo de la neurona haciendo referencia a las sucesiones de tiempos previas. Como se anticipó, si S no está elevando a ningún número, significa que hubo sólo una sucesión en el tiempo. Si está elevado a la 2, hubo dos sucesiones de tiempo, y de manera general, si está elevado a la n, hubo n sucesiones de tiempo.

Este correlato espacial expresado en las figuras 2 y 3 de las proposiciones temporales se justifica, de alguna manera, posteriormente en el teorema 6. Este nos indica que la facilitación y la suma temporal pueden ser reemplazadas por la suma espacial. Tanto la facilitación que remite a la amplitud en la recepción de señales neuronales, como la suma espacial y temporal que remiten la suma de las señales previas en cuanto a posición y tiempo, producen alteraciones en los ciclos. Sin embargo, las neuronas de McCulloch-Pitts abstraen estos cambios "factuales" y los autores conciben a sus redes invariantes respecto a estas alteraciones con motivos de simplificación representacional (McCulloch y Pitts 1943/1990, p. 101).

Como observación general, decimos que McCulloch y Pitts utilizan un lenguaje proposicional semejante al álgebra booleana para indicar la manera en que las neuronas lógicas se comunican, y un lenguaje de predicados para dar cuenta del comportamiento temporal de las neuronas. El modo de exposición de las ideas es semejante a la presentación de un cálculo lógico: detallan los componentes de su lenguaje, dan definiciones recursivas y demuestran teoremas. Sin embargo para algunos autores no se trataría de un cálculo en sentido estricto ya que un "sistema deductivo está destinado a capturar inferencias lógicas entre la clase de expresiones permitidas por la gramática del cálculo" (Piccinini, 2004,

p. 189). Si bien se demuestran teoremas no hay un conjunto de reglas sintácticas o reglas de inferencias a partir de las cuales se lleven a cabo las demostraciones. Otra de las críticas que se ha hecho al uso de la lógica apareció en una reseña temprana del artículo escrita por Frederic Fitch (1944):

> La lógica simbólica se usa con libertad e, incluso, con bastante descuido. La notación es innecesariamente complicada y el número de errores tipográficos es tan grande que el revisor no pudo descifrar varias partes cruciales del artículo. En algunos lugares los errores parecen ser de naturaleza lógica en lugar de meramente tipográficos, por ejemplo, la derivación de la fórmula (3) en la página 125. En cualquier caso, no existe una construcción rigurosa de un "cálculo lógico". (p. 49)

Sin embargo, a pesar de estas críticas, el lenguaje ha sido usado como un puntapié para desarrollar lenguajes de diseño lógico y redes neuronales, como veremos en el siguiente apartado.

6. Los efectos del modelo

Para Piccinini (2004) uno de los aportes del modelo de McCulloch y Pitts ha sido contribuir al desarrollo de una técnica para el diseño lógico de computadoras. El trabajo de 1943 "encuentra simpatía en una audiencia de científicos interesados en la epistemología pero entrenados en matemática o ingeniería más que en neurofisiología, como Norbert Wiener y John von Neumann." (Piccinini, 2003, p. 62). Esto se expresa en las Conferencias Macy y el Simposio Hyxon (1948), de las cuales surge la idea de que el estudio del cerebro y de las computadoras pertenece a la misma ciencia (Piccinini, 2003) y de esta manera se posibilita el desarrollo de distintos tipos de estudios:

1. La teoría de los autómatas de von Neumann.
2. El diseño de redes específicas usando las neuronas idealizadas para
3. funciones mentales específicas (ej.: modelo de la visión de McCulloch y Pitts de 1947 y el Perceptron de Rosenblatt de 1958[20]. Esto dio origen al desarrollo de estudios de redes neuronales.
4. El desarrollo de programas de computadora que ejecutan funciones > mentales
5. La neurofisiología experimental.[21] (Piccinini, 2003, p.110)

Los debates planteados por el artículo de 1943 estuvieron muy presentes en las Conferencias Macy en las que se llevaron a cabo importantes desarrollos teóricos para la investigación en IA. Es por ello que puede considerarse que "el modelo McCulloch-Pitts no engendra una notable posteridad ni en la lógica ni en la neurofisiología pero sí en la inteligencia artificial" (Dupuy, 2000, p. 62). En este sentido, podría decirse que si el aporte que ha dejado el modelo McCulloch-Pitts ha sido principalmente en el campo del diseño lógico y la teoría de los autómatas, el alcance representativo del modelo excedió la mente-cerebro de los autómatas naturales. El modelo produjo la posibilidad, o fue uno de los

[20] Si bien el Perceptron es un tipo particular del modelo neuronal McCulloch-Pitts, Rosenblatt, su creador, se demarcó de dichos autores por considerar su enfoque insuficientemente biológico (ver Dupuy, 2000, p. 62)

[21] Quizás la influencia más directa pueda ser el caso de Jerome Lettvin, tal y como comenta Piccinini: "El principal trabajo de Lettvin fue en este campo. Con Pitts, McCulloch y Humberto Maturana, Lettvin interpretó las señales que viajaban a través del nervio óptico de una rana como si fuera una implementación de los mecanismos descritos en el *Pandemonium* de Selfridge (Lettvin, Maturana, McCulloch y Pitts 1959). Su trabajo fue un gran avance en la interpretación de las propiedades de respuesta de las neuronas y se convirtió en un modelo para muchos trabajos en neurofisiología durante las décadas posteriores." (Piccinini, 2003, p. 111)

aportes claves en el diseño de autómatas artificiales, lo que muestra que un modelo puede producir un fenómeno *target* no esperado por los autores del mismo.

7. La aparición de las neuronas lógicas

Si bien consideramos que el *target* del modelo McCulloch Pitts es el fenómeno de la mente-cerebro, aparecen varias entidades en juego en dicha propuesta. Si hacemos un repaso, en el artículo de 1943 encontramos cuatro tipos de entidades conceptuales: las neuronas biológicas y sus redes, los psicones como átomos mentales y soporte de los procesos mentales, las proposiciones y, por último, las *neuronas lógicas o formales* y sus redes, producto de la equivalencia formal.

Donald Perkel (1988), en un artículo que repasa el legado teórico de McCulloch, llama a las neuronas McCulloch-Pitts "neuronas lógicas" para remarcar el aspecto formal de las mismas. En este sentido algunos autores han hablado de "neuronas formales" (Dupuy, 2000) o redes neuronales formales (von Neumann, 1951). Los propios creadores del modelo hablaron del carácter formal del mismo cuando distinguieron entre "equivalencia factual" y "equivalencia formal", distinción que tematizaremos en el próximo apartado.

También apelaron a la idea de "redes ficticias" para hablar de la substitución de redes con alteraciones por otras que no las tuvieran: "para redes que experimentan ambas alteraciones, podemos sustituir redes ficticias equivalentes compuestas de neuronas cuyas conexiones y límites son alterados" (McCulloch & Pitts, 1943/1990, p. 101). Con "alteraciones" se referían a aquellas cuestiones que modificaban toda la red: por un lado, a la facilitación y la extinción de las neuronas en las que la información temporalmente antecedente afecta la reacción posterior; y por otro lado, el aprendizaje que posibilita que un estímulo que anteriormente fue inadecuado deje de serlo. Con ficticias entonces se refieren en este caso a la idealización de las redes para simplificar su mecanismo. Tal como rastrea Dupuy (2000) en el recorrido teórico de McCulloch: si apelaba a la idea de "ficticias" era para referirse a las entidades idealizadas. Sin embargo, de acuerdo a la revisión que hace McCulloch del *paper* de 1943 a mediados de los años '60, la idealización de las neuronas no significaba que no se tratara de neuronas físicas. En este contexto el neurofisiólogo reniega de cierto devenir "desencarnado" e "incorpóreo" de la recepción de su *paper* en la teoría de los autómatas, que se presenta como un contrapunto con las posiciones de von Neumann y el uso que hace del lenguaje McCulloch-Pitts para el diseño de los primeros autómatas artificiales. (pp. 59, 61). Entonces el hecho de que los autores consideren a las redes como ficticias puede entenderse como una postura ficcionalista que considera a un modelo, en tanto idealización, como una ficción, en un marco que considera a la actividad misma de producir modelos como ficcional. O también puede entenderse que "ficticias" refiere a que tratan con entidades ficcionales: las neuronas lógicas o formales.

Sistema nervioso como conjunto de neuronas (*neuronas biológicas* y sus redes neuronales).	*Psicones* como átomos mentales. Las neuronas como soporte de la mente y por ende de los procesos mentales como el aprendizaje.	*Proposiciones* de la lógica proposicional bivalente.	Equivalencia formal: *neuronas lógicas* y sus redes.

Tabla 1: Tipo de entidades involucradas en el modelo de Mculloch y Pitts (1943/1990)

Entonces, si nos centramos en la recepción del lenguaje de McCulloch y Pitts como posibilitador del desarrollo de la teoría de los autómatas, podemos pensar lo siguiente: que, o bien las neuronas lógicas surgieron como representación de la mente-cerebro de los autómatas naturales y que, a su vez, posibilitaron la existencia de otro *target*, los autómatas artificiales; o bien, que los autómatas artificiales también son un modelo de la mente-cerebro humana. La respuesta a esta pregunta dependerá de nuestro vínculo con el humanismo y la concepción de la relación humano-máquina, lo cual no es debate de este artículo.

Lo que sí está claro es que los desarrollos teóricos de McCulloch y Pitts contribuyeron con gran protagonismo al despliegue de las investigaciones en inteligencia artificial. En este sentido, el fenómeno *target* del modelo parecería ser las neuronas lógicas. Y es en este sentido, que las neuronas lógicas en tanto objetos conceptuales no son triviales: implicaron un proceso de construcción y el desarrollo de nuevas áreas de investigación.

8. La equivalencia formal

La equivalencia formal es una de las cuestiones centrales del texto. Aparece incluso como aspecto clave en el *abstract* mismo, en donde se plantea como un objetivo del trabajo mostrar que

> Muchas opciones particulares entre los posibles supuestos neurofisiológicos son equivalentes, en el sentido de que por cada red que se comporta bajo una suposición, existe otra red que se comporta de acuerdo con la otra y da los mismos resultados, aunque quizás no al mismo tiempo. (McCulloch & Pitts, 1943/1990, p. 99)

Ya en el texto, la cuestión de la equivalencia formal es introducida luego de la presentación de dos posibles explicaciones acerca de la inhibición. Al respecto nos dicen que: "Dado que nos interesan las propiedades de las redes que son invariantes bajo equivalencia, podemos hacer las suposiciones físicas que sean más convenientes para el cálculo" (McCulloch & Pitts, 1943/1990, p. 100). Es decir, que si bien las suposiciones físicas se toman como punto de inspiración para el cálculo, no hay una pretensión de que coincidan exactamente sino que se pueden modificar en función de las necesidades de aquel, lo que explicamos en el apartado anterior cuando nos referimos al carácter ficticio de las redes.

McCulloch y Pitts distinguen la equivalencia formal de la explicación o equivalencia factual. En este contexto, aparece cierta ambigüedad en sus ideas, sin embargo, todas apuntan a una noción de identidad por correspondencia o identificación que, aunque no sea puesta explícitamente en términos de funciones biyectivas o isomorfismos, nos recuerda mucho a dichas ideas. ¿Qué significa la *equivalencia formal*? Al respecto aparecen tres ideas que nos parece que ayudan a vislumbrar la cuestión:

> 1. Las relaciones psicológicas que existen entre la actividad nerviosa *corresponden* a la relación entre proposiciones. La utilidad de la representación dependerá de la *identidad* de esas relaciones con la lógica de proposiciones;
> 2. Para cada reacción de una neurona hay una aserción *correspondiente* de una proposición -e implicará otra proposición-.
> 3. Las alteraciones como la facilitación, extinción y aprendizaje no afectan a las conclusiones que se saquen del *tratamiento formal* de la actividad de las redes nerviosas. (McCulloch & Pitts, 1943/1990, pp. 100–101)

Si tenemos en cuenta lo anteriormente dicho, aparecen varios sentidos respecto de la equivalencia formal, que se asocia a las nociones de identidad y correspondencia. En primer lugar, se habla de la correspondencia entre las proposiciones y las relaciones psicológicas que suponen en la actividad nerviosa. En segundo lugar, la correspondencia o identidad se da entre la lógica de proposiciones y las reacciones de las neuronas. Y, en tercer lugar, se remarca la relevancia del tratamiento formal y se aclara que las alteraciones (facilitación, extinción y aprendizaje) que implicarían un cambio en la información recibida por las neuronas y que afectarían a la red en su totalidad, no son tenidas en cuenta. Esto se debió a la necesidad de simplificar el cálculo. A su vez, se dice que la utilidad de la representación dependerá de la identidad entre las relaciones psicológicas o la actividad nerviosa de las neuronas y las proposiciones. De esta manera, la equivalencia formal es la garantía para que el formalismo sea una representación de la actividad neuronal y, en consecuencia, de los procesos psicológicos humanos soportados en los átomos mentales o psicones. De este modo se constituye en la explicación científica que garantiza el mecanismo de la mente-cerebro pero al mismo tiempo, en un criterio de isomorfismo del modelo, es decir, la correspondencia uno-a-uno entre el modelo y el *target* (Potochnik, Colombo, & Wright, 2019, p. 117).

A su vez, se dice que la utilidad de la representación dependerá de la identidad entre: relaciones psicológicas y la actividad nerviosa con las proposiciones, lo cual puede representarse según el siguiente esquema:

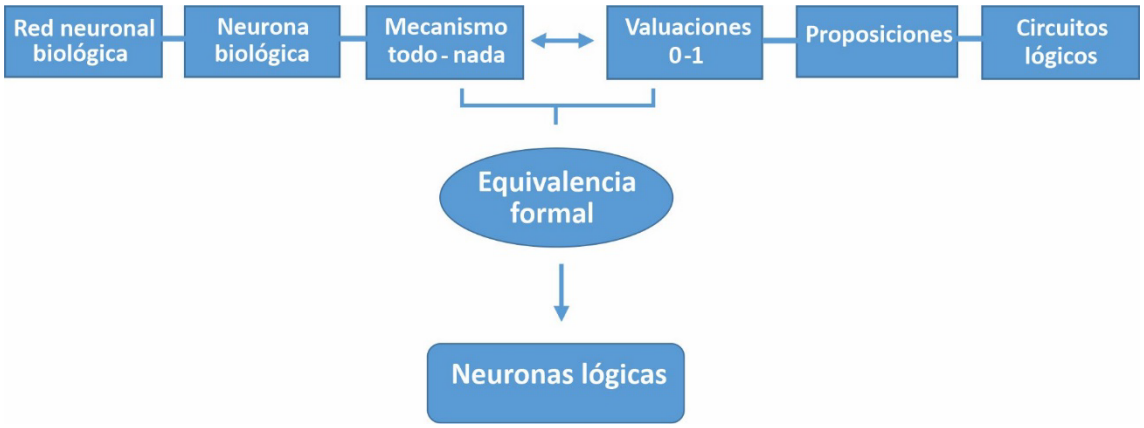

Figura 4.

De esta manera, la equivalencia formal puede entenderse ya no como el criterio de isomorfismo según el cual las neuronas lógicas representarían el *target* dado por la mente-cerebro, sino como el isomorfismo entre neuronas biológicas, psicones y proposiciones que producen una nueva entidad conceptual: las neuronas lógicas. La correspondencia entre las neuronas biológicas y el comportamiento de la valuación de las proposiciones lógicas, daría lugar a un *entre* que toma aspectos del comportamiento de ambas partes. Hemos visto que en este marco la equivalencia formal ha cumplido un rol fundamental: en tanto explicación científica, en tanto criterio de isomorfismo entre el modelo y el fenómeno target y finalmente como criterio de identidad de las entidades conceptuales producidas en el marco del modelo: las neuronas lógicas.

9. Conclusión

En este trabajo hemos visto que el desarrollo del modelo neuronal de McCulloch y Pitts ha producido varios aportes los cuales retomamos, desarrollamos y, a partir de ellos, analizamos las implicancias de dicho modelo en la historia de la ciencia. En primer lugar, en tanto teoría computacional de la mente hemos visto que esta se debió a la aplicación de la noción de computación de Turing. Esta teoría de la mente se trató de una explicación mecanicista que se insertó en los desarrollos de la neurofisiología teórica de Ramón y Cajal, Barenne, entre otros. La forma en que se llevó a cabo dicha teoría fue a partir de la resolución del problema mente-cuerpo asociando una teoría de la mente que venía desarrollando McCulloch —la teoría de los psicones— a la teoría de la sinapsis neuronal. A su vez estas unidades nerviosas-mentales fueron asociadas a proposiciones lógicas a partir de la correspondencia establecida entre el mecanismo todo-nada de las neuronas y la lógica bivalente del álgebra de Boole. A su vez, influenciados por los estudios de biología matemática de Woodger, Rashevsky y Householder desarrollaron modelos de redes neuronales a través de la producción de un lenguaje lógico, para el cual fueron también influencias importantes Russell y Carnap.

Por otro lado, el uso del lenguaje II de Carnap y la lectura de los *Principia Mathemathica* posibilitaron la creación de una escritura para indicar información temporal al *interior* de las proposiciones. O, dicho de otra manera, para incluir en el *contenido* de las proposiciones, la dimensión temporal de la comunicación de las neuronas lógicas. A cada fórmula con su operación, le correspondía una imagen del tipo presentado en el apartado anterior. A su vez, la influencia de Russell-Whitehead y Carnap permitió organizar el texto en forma de cálculo con definiciones recursivas, deducción de teoremas y el uso de la inducción matemática como forma de generalizar. Sin embargo esta lectura tiene un contrapunto en la crítica de Fitch y Piccinini acerca de la necesidad de reglas sintácticas para considerar a un lenguaje un cálculo lógico.

La confluencia de distintos leguajes lógico-matemáticos con algunas ideas de la neurofisiología produjo un nuevo lenguaje que posibilitó la existencia de un nuevo tipo de neuronas. En particular, la equivalencia formal del mecanismo todo-nada de la neurona biológica con la lógica binaria (0-1) del álgebra de Boole. Pero no sólo la correspondencia en el mecanismo de una lógica bivalente posibilitó la construcción de un nuevo lenguaje. Pareciera que también la semejanza en la representación visual de las neuronas biológicas y los diagramas de flujo de las compuertas OR, AND y AND-NOT. La *yuxtaposición* de la representación gráfica de la neurona biológica con los diagramas posibilitó una escritura espacial que sintetizó dos representaciones en una nueva. Estos nuevos objetos no eran exactamente iguales a las neuronas biológicas ni exactamente iguales a la representación gráfica de las compuertas lógicas de Boole.

En este sentido, la equivalencia formal, que tuvo inicialmente la motivación de explicar los procesos mentales, produjo una nueva entidad conceptual: las neuronas lógicas, que a su vez posibilitó el desarrollo de nuevas investigaciones científicas. Es así que posibilitaron en la historia de la cibernética el paso del uso de la noción de computación para entender la mente al uso de la noción de mente para diseñar mecanismos computacionales.

En tanto modelo científico que pretende representar el mecanismo de la mente-cerebro, podría pensarse que el lenguaje de McCulloch y Pitts lleva a cabo dicha representación. Pero también podría pensarse que, en la creación del *target* del modelo, se

amplía el alcance representativo de éste último. Ya no sólo representa la mente-cerebro sino también la posibilidad del desarrollo de máquinas inteligentes. Para ello fue necesario que la equivalencia formal funcione como explicación científica en contraposición a la explicación factual y que haya una concepción mecanicista de dicha equivalencia según la cual mente, cerebro y mecanismo convergen. Podemos pensar, entonces, que lo que ahora es recordado como el modelo neuronal McCulloch-Pitts es producto de conjugar diferentes disciplinas de tradiciones formales y experimentales.

Referencias

Abraham, T. H. (2002). (Physio)Logical Circuits: The intellectual origins of the McCulloch-Pitts Neural Networks. *Journal of the History of the Behavioral Sciences*, 38 (1), 3–25. https://doi.org/10.1002/jhbs.1094

Arbib, M. (1987). *Brains, machines and mathematics*. Springer Verlag.

Barberis, S. (en prensa). La explicación mecanicista. En prensa.

Branca, M. I., Ramírez, A. O., y Vilatta, M. E. (2015). Modelos de explicación en psicología cognitiva y neurociencias. *Anuario de Investigaciones de la Facultad de Psicología*, 2 (1), 176-191. https://www.revistas.unc.edu.ar/index.php/aifp

Boolos, G., Burgess, J., Jeffrey, R. (2007). *Computability and Logic*. Cambridge University Press.

Carnap, R. ([1937] 2007). *Logical Syntax of Language*. Routledge.

Cassini, A. (2018). Models without a Target. *Artefactos. Revista de estudios de la ciencia y la tecnología*. 7 (2), 185-209. http://dx.doi.org/10.14201/art201872185209

Craver, C. (2007). *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience.* Oxford, Clarendon Press.

Dupuy, J-P. (2000). *The mechanization of the mind: On the origins of cognitive science*. Princeton University Press.

Fitch, F. (1944). Review of McCulloch and Pitts 1943. *Journal of Symbolic Logic*. 9 (2), 49–50. https://doi.org/10.2307/2268029

Householder, A. (1941a). A theory of steady-state activity in nerve-fiber networks: I. Definitions and preliminary lemmas. *Bulletin of Mathematical Biophysics*, 3, 63–69. https://doi.org/10.1007/BF02478220

Householder, A. (1941b). A theory of steady-state activity in nerve-fiber networks II: The simple circuit. *Bulletin of Mathematical Biophysics*, 3, 105–112. https://doi.org/10.1007/BF02478168

Householder, A. (1941c). A theory of steady-state activity in nerve-fiber networks III: The simple circuit in complete activity. *Bulletin of Mathematical Biophysics*, 3, 137–140. https://doi.org/10.1007/BF02477933

Householder, A. (1942). A theory of steady-state activity in nerve-fiber networks IV: N circuits with a common synapse. *Bulletin of Mathematical Biophysics*, 4, 7–14. https://doi.org/10.1007/BF02477350

Kleene, S. C. (1956), Representation of Events in Nerve Nets and Finite Automata en C. E. Shannon y J. McCarthy (Eds.), *Automata Studies*, (3-42). Princeton University Press.

Lawvere, F. W. & Schanuel, S. (1997). *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press.

(McCulloch & Pitts, 1943) McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115-133. https://doi.org/10/djsbj6

McCulloch, W. y Pitts, W. (1990). A logical calculus of de ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 52 (l/2), 99-115. (Obra original de 1943) https://doi.org/10.1007/BF02459570

Northrop, F. S. C. (1931). *Science and first principles*. Macmillan.

Perkel, D. (1988). Logical neurons: the enigmatic legacy of Warren McCulloch. *Trends in Neurosciences*, 11 (1), 9-12. https://doi.org/10.1016/0166-2236(88)90041-0

Piccinini, G. (2003). *Computations and computers in the sciences of mind and brain*. [Tesis de doctorado, Universidad de Pittsburgh]. https://www.researchgate.net/publication/301340375_Computation_and_Computers_in_the_Sciences_of_Mind_and_Brain

Piccinini, G. (2004). The first computational theory of mind and brain: a close look at McCulloch and Pitts' "Logical calculus of de ideas immanent in nervous activity". *Synthese*, 141, 175–215. https://doi.org/10.1023/B:SYNT.0000043018.52445.3e

Potochnik, A., Colombo, M., Wright, C. (2019). *Recipes for Science. An Introduction to Scientific Methods and Reasoning*. Editorial Routledge.

Schlatter, M. y Aizawa, K. (2008). Walter Pitts and "A Logical Calculus". *Synthese*, 162, 235–250. https://doi.org/10.1007/s11229-007-9182-9

von Neumann, J. (1945). *First Draft of a Report on the EDVAC*. Technical Report, Moore School of Electrical Engineering, University of Pennsylvania. En línea: http://abelgo.cn/cs101/papers/Neumann.pdf

von Neumann, J. (1951). The General and Logical Theory of Automata. En L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 1–41). Wiley.

Turing, A. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*(1), 230-265. En línea: https://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf

Zamora-Cárdenas, W., Zumbado, M., Trejos-Zelaya (2020). McCulloch-Pitts Artificial Neuron and Rosenblatt's Perceptron: An abstract specification in Z. *Revista Technology inside by CPIC*, 5 (5), 16-29. https://cpic-sistemas.or.cr/revista/index.php/technology-inside/article/view/36/27

Declaraciones

Conflictos de interés: La autora declara que no existen conflictos de interés.