Capítulo 2

Viejos principios y nuevas aplicaciones: aprendizaje y reforzamiento en inteligencia artificial

Álvaro Torres-Chávez[1] *y Ángel Eugenio Tovar*
Universidad Nacional Autónoma de México

Resumen

Describimos tres aportes fundacionales de la psicología experimental relacionados con los términos Aprendizaje y Reforzamiento, indicamos sus distintos significados, complementarios desde las aportaciones de Thorndike, Skinner y Hebb. Después, inspirados en la propuesta de Sutton y Barto, indicamos las relaciones de estos principios y conceptos con la inteligencia artificial y el Aprendizaje de Máquinas contemporáneo, en particular con los modelos basados en Aprendizaje por Reforzamiento. Sugerimos que la investigación psicológica actual se puede beneficiar de la necesidad de formalización

1 Correspondencia: Dirigirse a Álvaro Torres-Chávez (alvarot@unam.mx) y Ángel Tovar (aetovar@unam.mx)

conceptual, y de las posibilidades de implementación que permite el modelamiento computacional.

Palabras clave: aprendizaje por reforzamiento, redes neuronales, algoritmos de aprendizaje, inteligencia artificial, ChatGPT.

Abstract

Here we describe three foundational psychological contributions related to the concepts of Learning and Reinforcement. We describe their different but complementary meanings from the perspectives of Thorndike, Skinner and Hebb. We then, inspired by the approach of Sutton and Barto, describe the relationship between these concepts and the fields of contemporary Artificial Intelligence and Machine Learning, particularly with those models based on Reinforcement Learning. We finally suggest that current developments in psychological research can benefit from conceptual formalization and computational implementations.

Keywords: Reinforcement Learning, Neural Networks, Learning Algorithms, Artificial Intelligence, GPT Chat.

La Psicología Experimental y su perspectiva Analítico-Conductual han tenido dos términos básicos a lo largo de su historia: Aprendizaje y Reforzamiento. A pesar de su relevancia dentro del diseño de estrategias de investigación e intervención conductual, tales términos han sido polisémicos, y con alcances e implementaciones variables a nivel metodológico. En este capítulo discutimos que, pese a dicha variabilidad, el desarrollo teórico y conceptual de estos términos ha crecido hasta volverse una de las principales fuentes para las actuales aplicaciones en inteligencia artificial (Doya, 2023), como es el caso del ChatGPT-3 (Generative Pre-trained Transformer), desarrollado como un modelo de lenguaje (OpenAI, 2023). Enseguida presentamos las leyes o principios que postularon tres de los investigadores históricamente claves en Psicología Experimental: Edward L. Thorndike, Burrus F. Skinner y Donald O. Hebb, en torno al Aprendizaje y el Reforzamiento, y discutimos cómo sus conceptualizaciones influyen contemporáneamente en el "aprendizaje por reforzamiento" y las "redes neuronales artificiales" usadas en inteligencia artificial. Haremos

un "viaje" entre los viejos principios psicológicos y sus aplicaciones actuales en inteligencia artificial.

Parte 1. Thorndike y la identificación de principios

Si rastreamos el uso del término Aprendizaje, con sus implicaciones contemporáneas en la Psicología Científica (i.e., experimental), podemos encontrar referentes destacados en trabajos de Edward Thorndike (1898, 1911, 1927, 1931), donde postuló la denominada "Ley del Efecto" y describió el "Aprendizaje por Ensayo y Error". Se ha vuelto un lugar común decir que la Ley del Efecto se circunscribe al enunciado:

> De las numerosas respuestas dadas en una situación, aquellas acompañadas o seguidas de satisfacción para el animal estarán ( ) más firmemente conectadas con la situación, de tal forma que al repetirse dicha situación, éstas respuestas serán más probables de recurrir ( ) Aquellas respuestas acompañadas o seguidas de incomodidad para el animal ( ) tendrán sus conexiones debilitadas con la situación, de tal forma que al repetirse la situación, estas respuestas serán menos probables. Cuanto mayor sea la satisfacción o la incomodidad, mayor será el fortalecimiento o debilitamiento del vínculo. (Thorndike, 1911, p. 244).

El "Aprendizaje" desde la perspectiva de Thorndike es el establecimiento de relaciones, vínculos, conexiones o asociaciones entre eventos (e.g., estímulos, respuestas) en situaciones específicas. Una revisitación de la publicación "Animal Intelligence. Experimental Studies" de Thorndike (1911), nos permite recordar que su perspectiva fue más amplia. La Ley del Efecto se formuló junto con otros principios de la conducta en el contexto de "Leyes e Hipótesis para la Conducta" que abordaron, entre otros problemas: el aprendizaje asociativo; el establecimiento de vínculos entre situaciones y respuestas (S-R); la fuerza de esos vínculos, conexiones, relaciones o asociaciones S-R con base en su probabilidad ("likely"); y el sustrato neuronal y sináptico de la fuerza del vínculo S-R en términos de la eficacia sináptica o "neurone intimacy" como la llamó Thorndike.

Todos estos aspectos solo fueron mencionados por Thorndike, pero no los desarrolló; sin embargo, se convirtieron en centrales para los posteriores desarrollos en teorías de aprendizaje, cognición y neurociencia. Más aún, son

todos componentes centrales en la visión contemporánea que permite los desarrollos conceptuales y de implementación en inteligencia artificial (IA), desde el Aprendizaje Automático o de Máquinas ("Machine Learning"), el Aprendizaje por Reforzamiento (Reinforcement Learning) y el procesamiento en Redes Neuronales Artificiales (e. g., profundas), a partir principalmente de algoritmos de maximización de la recompensa, que parecen seguir los principios de la Ley del Efecto, mediante mecanismos de modificación de fuerzas asociativas que comentaremos más adelante.

Parte 2. Skinner, la función y la experiencia

Dentro los numerosos aportes de B.F. Skinner (1938, 1953, 1969, 1981), queremos destacar su trabajo teórico respecto al "Reforzamiento", desarrollado como la noción de probabilidad de respuesta. Skinner se refirió a la "frecuencia de respuesta" como el indicador de su probabilidad de ocurrencia, la cual está funcionalmente determinada por sus consecuencias. Propuso una situación experimental controlada para la observación e interpretación de las frecuencias, la conocida "Caja de Skinner". Este espacio de control y manipulación experimental le permitió observar de modo claro, y medir con precisión, los cambios producidos por diferentes consecuencias ambientales sobre la frecuencia de ocurrencia de respuestas específicas. Al criticar las "ficciones explicativas" tales como "el pichón adquirió el hábito de levantar su cabeza a cierta altura" (1953, p. 64), Skinner introdujo el importante concepto de "contingencia", es decir, una relación de dependencia probabilística entre la ocurrencia de una consecuencia dada la emisión de una respuesta. Los alcances del concepto de contingencia pueden observarse incluso en el desarrollo de algunas perspectivas Bayesianas (Griffiths *et al*., 2023) que son ampliamente usadas tanto en el desarrollo teórico de la psicología científica contemporánea, como en los modelos actuales de Aprendizaje por Reforzamiento de la IA (Doya, 2023; Russell & Norvig, 2022; Sutton & Barto, 2018).

Skinner (1953) desarrolló muchos conceptos técnicos tales como "discriminación operante", donde estímulos específicos llamados "discriminativos" constituyen la ocasión o situación en que la emisión de una respuesta hace más probable una consecuencia particular. Este proceso de control de estímulos, así como muchos otros de la perspectiva Skinneriana, ha tenido una gran

repercusión en el desarrollo teórico y de implementación computacional del Aprendizaje por Reforzamiento o reinforcement learning, desarrollado por Sutton & Barto (2014), según estos autores, en su teoría, hay una gran similitud entre control de estímulos y la "tarea de búsqueda asociativa" que consiste en un "mapeo" o identificación y establecimiento de relaciones entre situaciones y acciones que permiten maximizar el reforzamiento por agentes artificiales.

Por otra parte, el concepto de "conducta operante", central en la obra Skinneriana, describe una clase de respuestas que son semejantes entre sí y emitidas espontáneamente por el organismo, que pueden volver a ocurrir en un futuro y "operan" sobre el medioambiente para generar o hacer más probables ciertas consecuencias. Sutton & Barto (2017) reconocen explícitamente haber retomado ese concepto para su formulación del Aprendizaje por Reforzamiento:

> El aprendizaje por reforzamiento es aprender qué hacer –cómo mapear situaciones a acciones– para maximizar una señal de recompensa numérica. Al aprendiz no se le dice qué acciones tomar, sino que debe descubrir qué acciones producen la mayor recompensa al probarlas..., las acciones pueden afectar no solo la recompensa inmediata, sino también la siguiente situación y, a través de eso, todas las recompensas posteriores (Sutton & Barto, 2018, p. 1-2).

Estos investigadores acentúan que el aprendizaje se genera mediante las interacciones en el ambiente. Proponen que la manera de estudiarlo es diseñar modelos o máquinas computacionales efectivas para solucionar problemas de aprendizaje (mapeo) y evaluar sus diseños mediante el análisis matemático o la experimentación computacional. Los modelos de aprendizaje por reforzamiento abarcan el problema de un agente (o aprendiz) encaminado al logro de objetivos, interactuando a lo largo del tiempo en un entorno con incertidumbre.

Sutton y Barto (2018) identifican en su modelo de aprendizaje por reforzamiento seis elementos principales como sigue:

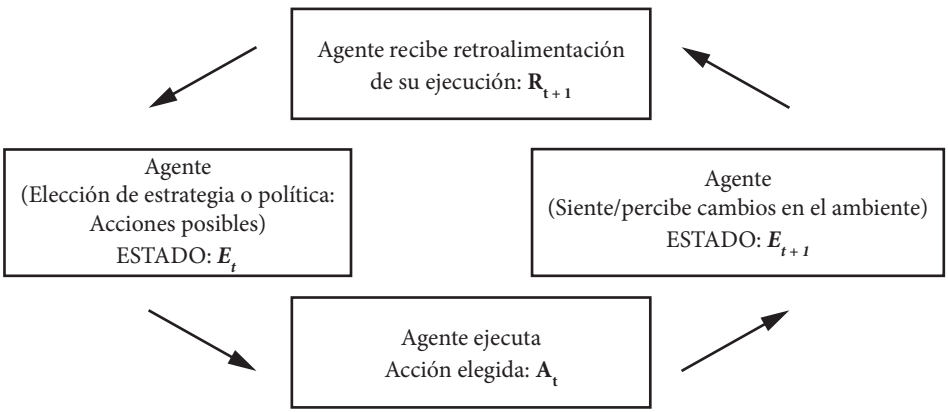
1. Un agente de aprendizaje (o aprendiz), que toma decisiones sobre cuáles acciones realizar para obtener recompensas.
2. Un ambiente es el entorno del agente, está constituido por "estados" o configuraciones discretas de estímulos percibidos por el agente; este componente se puede relacionar con los "affordances" (Gibson, 1977).

3. Una política o estrategia para elegir formas de comportarse en un momento dado (un mapeo de los estados percibidos del ambiente con las acciones a realizar cuando se encuentra nuevamente en esos estados).
4. Una señal de recompensa o reforzador, que determina la deseabilidad inmediata de los estados ambientales.
5. Una función de valor, que especifica la deseabilidad a largo plazo de los estados, luego de estimar el valor de los posibles estados que sigan a futuro; así como las recompensas disponibles en esos estados.
6. Un modelo del ambiente de aprendizaje (elemento opcional), es decir, una representación cognitiva de los estímulos, su organización y función que permite hacer inferencias, *planes* y predicciones sobre el efecto potencial de acciones sobre el futuro. Los métodos que incluyen este elemento son considerados "basados en modelos" y son opuestos a los métodos "libres de modelos", que son explícitamente aprendices por ensayo y error.

En la Figura 1 ilustramos los componentes de este modelo de aprendizaje por reforzamiento y algunas de sus relaciones.

Figura 1

Interacción Agente-Ambiente en modelos de Aprendizaje por Reforzamiento



Los valores de recompensa estimados y obtenidos determinan las políticas, si una acción elegida por una política obtiene baja recompensa, entonces la política debe cambiarse para seleccionar alguna otra acción futura bajo esa misma situación.

Sutton y Barto (2017) indican que las señales de recompensa pueden ser funciones estocásticas del estado del ambiente y de las acciones tomadas. Una función estocástica es aquella en la cual los resultados son en parte aleatorios y en parte bajo control del agente que toma decisiones.

En síntesis, los modelos computacionales del aprendizaje por reforzamiento derivados de la perspectiva de Sutton y Barto, simulan procesos de toma decisiones de un agente, según este percibe (sense) las características y rasgos de su ambiente, para elegir qué acciones puede o debe emitir, y luego el agente observa las consecuencias y les otorga un valor de recompensa (Doya, 2023). En los primeros episodios (e.g., ensayos) la elección es "a ciegas" y principalmente aleatoria, se ensayan acciones y se observa su efecto para cambiar el entorno y se valoran las recompensas obtenidas.

Las decisiones secuenciales tomadas por el aprendiz se modelan matemáticamente, esto puede ser mediante numerosos métodos, por ejemplo, con procesos de Decisión de Markov, que consideran estados del ambiente, acciones y transiciones probabilísticas entre estados. Cada acción tiene una probabilidad asociada de llevar al agente a un nuevo estado y se asocia con una recompensa en cada transición de estados.

Por otra parte, un caso ejemplar de un método "libre de modelos" es el algoritmo "Q-learning" (Quality Learning), que no toma en cuenta las probabilidades de transición, y es independiente a políticas de comportamiento, pero permite aprender una estrategia óptima, la función Q representa el valor esperado acumulado de una acción en un estado dado. Q-Learning se puede usar para entrenar redes neuronales artificiales, en función de las recompensas y los estados observados durante la interacción del agente con el entorno, y esto puede ayudar a maximizar las recompensas a largo plazo. La expresión matemática del algoritmo Q-learning[2] es:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha\Big[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\Big]. \quad (1)$$

2 En la Ecuación 1 se muestra el Algoritmo Q-Learning control de diferencias temporales. (citado en Sutton & Barto, 2017, p. 131).

El valor Q se actualiza sumando el valor Q anterior más una tasa de aprendizaje (alfa) que multiplica al valor Q de la acción seleccionada, el cual consiste de la recompensa obtenida más un factor de descuento (gamma) multiplicado por la estimación del valor futuro óptimo. Los valores Q son estimaciones o predicciones de las recompensas, y las decisiones del agente inteligente respecto a qué acciones realizar se basan en sus juicios de valor, es decir, en sus cálculos acerca de las recompensas futuras con base en sus experiencias anteriores.

Como se pudo observar en esta sección, los trabajos en *reinforcement learning*, destacadamente lo elaborado y conjuntado por Sutton y Barto (1998, 2018) han desarrollado implementaciones computacionales para analizar y explicar el proceso aprendizaje de un agente que se basa en, y a la vez amplía, conceptos y métodos de la conducta operante formulados por B. F. Skinner. Esta ejemplificación es tan solo una breve revisión de uno de los aportes impulsados por Sutton y Barto al creciente campo de algoritmos e implementaciones en inteligencia artificial.

Es importante notar que, aunque la explicación de Sutton y Barto se denomina "aprendizaje por reforzamiento", tiene una perspectiva más amplia que incluye aspectos más allá de los conductuales tradicionales, que puede considerarse cognitiva, en tanto considera procesos de elección y toma de decisiones para la predicción o estimación de valor estado-acción.

Presentaremos ahora una tercera aportación clave para la conceptualización del aprendizaje y el reforzamiento.

Parte 3. Hebb y los mecanismos

Otro de los principios psicológicos que ha tenido un gran impacto en la investigación conductual, en neurociencias y en la inteligencia artificial es la llamada "Regla de Aprendizaje de Hebb", fue formulada por Donald O. Hebb en su libro *The Organization of Behavior. A Neuropsychologial Theory* (1949).

Hebb presentó su "postulado neurofisiológico" que, en línea con las ideas de Thorndike, describe los efectos de las relaciones entre neuronas de la siguiente forma:

> Supongamos que la persistencia o repetición de una actividad reverberante (o "huella") tiende a inducir cambios celulares duraderos que se suman a su

> estabilidad. La suposición se puede establecer con precisión de la siguiente manera: cuando un axón de la célula A está lo suficientemente cerca como para excitar una célula B y participa repetida o persistentemente en su activación, se produce algún proceso de crecimiento o cambio metabólico en una o ambas células, de tal forma que la eficiencia de A, como una de las células que hace disparar a B, aumenta. (Hebb, 1949, p. 62).

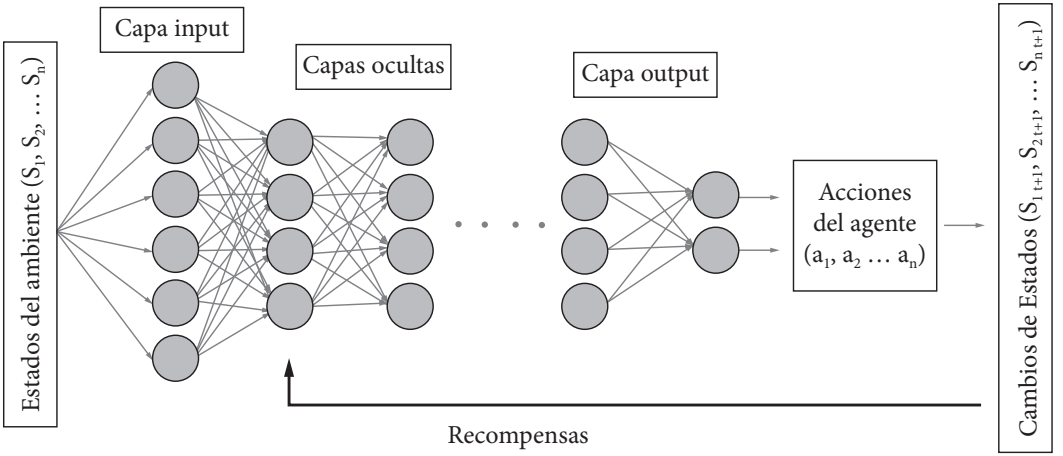
La Regla de Hebb, aunque inicialmente teórica, captura a su vez mecanismos neurofisilógicos de mejora en la eficiencia sináptica y procesos funcionales de almacenamiento de información. En el tiempo, esta regla se tradujo en algoritmos de aprendizaje en redes neuronales artificiales (ensambles, conjuntos o asambleas neuronales, en palabras de Hebb).

Las redes neuronales artificiales hacen referencia a una amplia familia de modelos que se pueden diseñar con numerosas arquitecturas y algoritmos. La arquitectura describe el número de neuronas (o unidades de procesamiento) y su arreglo y conectividad en capas. Los algoritmos describen cómo se propaga la activación a través de las neuronas, y cómo se ajustan las conexiones entre estas durante el proceso de aprendizaje. Existen redes sencillas que pueden ser monocapa, o de capa única, donde la activación simplemente reverbera en un arreglo de unidades que usualmente están "completamente conectadas" (o *fully connected*, conectadas todas con todas). Y existen redes "multicapa" donde la activación se calcula para un grupo de neuronas, antes de ser enviada al siguiente grupo de neuronas, siguiendo un flujo usualmente "hacia adelante" desde una capa de entrada (o estimulación) hasta una de salida (o respuesta). En psicología, las redes tricapa se popularizaron desde los años ochenta por su potencial para capturar y explicar, de manera sencilla, procesos cognitivos complejos, mediante la teoría conexionista (Thomas & McClelland, 2023).

Las redes neuronales profundas, de amplio uso en sistemas de inteligencia artificial contemporáneos, se denominan como tal por tener numerosas capas de procesamiento intermedias (i.e., más de 3 capas en total). Tienen al menos una capa de entrada, decenas o cientos de capas intermedias (denominadas como "ocultas"), y al menos una capa de salida. Veáse por ejemplo la Figura 2, que muestra una Red Neuronal Profunda integrada en un ambiente de Aprendizaje por Reforzamiento.

Figura 2

Red Neuronal Profunda integrada en un sistema de Aprendizaje por Reforzamiento.



El proceso de aprendizaje en una red neuronal artificial se logra cuando la actividad de sus neuronas y sus conexiones, análogas a sinapsis y modeladas numéricamente, convergen en un patrón de actividad que ***minimiza el error*** (o diferencia) entre los niveles de activación de los "outputs" actuales y los esperados. Si bien los algoritmos de aprendizaje de las redes tricapa y profundas frecuentemente se basan en el algoritmo de *retroprogapación* para la minimización del error (Rumelhart *et al*., 1986), el concepto de "Aprendizaje" en estas redes (Shultz, 2003) operacionalmente se basa en el fundamento Hebbiano de modificación de las conexiones (sinapsis) como base de la mejora en la eficiencia de la red.

Los principios psicológicos mencionados continúan influyendo la concepción y estudio del Aprendizaje y el Reforzamiento contemporáneos, con otros significados nuevos tanto en el campo de la Psicología como en las áreas de la inteligencia artificial, el Aprendizaje Automático o de Máquina y el Aprendizaje por Reforzamiento. Por ejemplo, dentro de las aplicaciones más frecuentes de la inteligencia artificial se encuentra el uso de redes neuronales profundas y sistemas de aprendizaje por reforzamiento para diagnóstico médico clínico. Tal es el caso en el análisis e identificación de patrones en imágenes de resonancia magnética funcional para detección de cáncer de mama (Sasaki *et*

al., 2020; Sechopoulos *et al*., 2021); para tratamientos psicológicos, e.g., en el diseño y aplicación de programas de terapia asistida por computadora que ayudan a que los pacientes aprendan y practiquen habilidades de afrontamiento y autocontrol (Fitzpatrick *et al*., 2017); y por supuesto para el análisis del lenguaje natural, la traducción automática y la generación de texto como el ChatGPT, entre muchas otras aplicaciones científicas, comerciales e industriales (OpenAI, 2023). En el ámbito de la investigación académica, donde de hecho se han originado los conocimientos básicos que sustentan las aplicaciones modernas, también existen usos científicos para el desarrollo de metodologías y teorías modernas sobre una amplia gama de fenómenos psicológicos, v.gr. el modelamiento computacional de la formación de conceptos, la categorización y conducta simbólica en general (Tovar *et al*., 2023).

Consideraciones finales

Thorndike describió mediante la investigación experimental que la conducta está determinada primordialmente por sus consecuencias, el agente de Sutton y Barto emite/ensaya acciones/estrategias que son "seleccionadas" por el valor inmediato o a largo plazo que les estima el agente y que le permiten tomar decisiones en su búsqueda para maximizar la recompensa acumulada a lo largo de su interacción con un ambiente incierto.

Skinner profundizó los controles metodológicos del análisis experimental de la conducta. Describió el proceso de reforzamiento, es decir, la selección por sus consecuencias de respuestas emitidas espontáneamente; los eventos no tienen propiedades reforzantes inherentes, sus propiedades reforzantes se identifican a través del análisis funcional. Son reforzantes si se observan aumentos en la probabilidad de ocurrencia de la respuesta o son aversivos si hay disminución en la probabilidad (Skinner, 1981). En Sutton y Barto (2018), el agente participa activamente en el proceso de selección, estimando valores de reforzamiento estado-acción en función de su interacción con el ambiente; el aprendizaje se produce en la interacción misma.

Por otra parte, en las redes neuronales (recurrentes, profundas, convolucionales, de memoria de corto plazo extensa—LSTM, o acopladas con transformadores) el aprendizaje es en buena medida consecuencia de cambios en los pesos de conexiones entre neuronas artificiales, es decir, de modificaciones en eficacia

sináptica, tal como prescribe la Regla de Hebb y anticiparon algunas ideas de Thorndike. Otros modelos de redes neuronales dinámicas describen el aprendizaje como producto de modificaciones en la arquitectura de la red, mediante procesos semejantes a la neurogénesis y a la apoptosis; desarrollo y modificación de "ensambles" en términos Hebbianos (Shultz *et al*., 2023).

El conocimiento y análisis detallado de los principios computacionales más importantes de la inteligencia artificial y del Aprendizaje por Reforzamiento podrían ser de beneficio mutuo para la investigación del aprendizaje en las Ciencias Cognitivas y del Comportamiento, en particular para el Análisis Experimental de la Conducta, mediante el uso de sistemas y modelos artificiales o naturales. Nuestro llamado no es nuevo (véase Commons *et al*., 1991; Burgos, 2000), pero sí queremos subrayar que el vínculo analizado permite beneficios mutuos, pues las implementaciones computacionales pueden beneficiarse de las teorías conductuales y cognitivas de la psicología científica, y estas teorías se pueden evaluar, rechazar y desarrollar con herramientas computacionales que permiten su formalización, implementación y análisis.

Referencias

Burgos, J. E. (2000). Neural Networks in Behavior Analysis: Models, Results, and Issues. *Mexican Journal of Behavior Analysis*, 26, 129-134.

Commons, M. L., Grossberg, S., & Staddon, J. E. R. (Eds.). (1991). *Neural network models of conditioning and action.* Lawrence Erlbaum Associates, Inc.

Doya, K. (2023). Reinforcement Learning. En R. Sun (Ed.) *The Cambridge Handbook of Computational Cognitive Sciences.* (2nd ed., 350-369). Cambridge University Press.

Gibson, J. J. (1977). The Theory of Affordances. En R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology. (pp.* 67-82). Erlbaum.

Griffiths, Th. L., Kemp, Ch., & Tenenbaum, J. B. (2023). Bayesian Models of Cognition. En R. Sun (Ed.) *The Cambridge Handbook of Computational Cognitive Sciences.* (2nd ed., 80-138). Cambridge University Press.

Open AI. (2023). *Introducing chatGPT*. Recuperado de https://openai.com/blog/chatgpt

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation. En D. Rumelhart, McClelland, J. L. & PDP Research Group (Eds.). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 1 Foundations*. (pp. 318-362). MIT Press.

Russell, S. J. & Norvig, P. (2021). *Artificial Intelligence. A Modern Approach*. (4th Ed.). Prentice-Hall.

Sasaki, M., Tozaki, M., Rodríguez-Ruiz, A., Yotsumoto, D., Ichiki, Y., Terawaki, A., Oosako, S., Sagara, Y., & Sagara, Y. (2020). Artificial intelligence for breast cancer detection in mammography: experience of use of the ScreenPoint Medical Transpara system in 310 Japanese women. *Breast cancer (Tokyo, Japan)*, *27*(4), 642–651. https://doi.org/10.1007/s12282-020-01061-8

Sechopoulos, I., Teuwen, J., & Mann, R. (2021). Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Seminars in cancer biology*, *72*, 214–225. https://doi.org/10.1016/j.semcancer.2020.06.002

Shultz, Th. R. (2003). *Computational Developmental Psychology*. MIT Press.

Shultz, Th. R. & Nobandegani, A. S. (2023). Computational Models of Developmental Psychology. En R. Sun (Ed.) *The Cambridge Handbook of Computational Cognitive Sciences.* (2nd ed., 769-794). Cambridge University Press.

Skinner, B. F. (1938). *The Behavior of Organisms. An Experimental Analysis*. Appleton Century-Crofts.

Skinner, B. F. (1953). *Science and Human Behavior*. Macmillan.

Skinner, B. F. (1969). *Contingencies of Reinforcement. A Theoretical Analysis***.** Appleton Century-Crofts.

Skinner B. F. (1981). Selection by consequences. *Science (New York, N.Y.)*, *213*(4507), 501–504. https://doi.org/10.1126/science.7244649

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. The MIT Press.

Thomas, M. S. C. & McClelland, J. L. (2023). Connectionist Models of Cognition. In R. Sun (Ed.) *The Cambridge Handbook of Computational Cognitive Sciences.* (2nd ed., 29-79). Cambridge University Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements, 2*(4), i–109. https://doi.org/10.1037/h0092987

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* Macmillan Press. https://doi.org/10.5962/bhl.title.55072

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology, 39,* 212–222. https://doi.org/10.2307/1415413

Thorndike, E. L. (1931). *Human learning*. The Century Co. https://doi.org/10.1037/11243-000

Tovar, Á. E., Torres-Chávez, Á., Mofrad, A. A., & Arntzen, E. (2023). Computational models of stimulus equivalence: An intersection for the study of symbolic behavior. *Journal of the Experimental Analysis of Behavior*, *119*(2), 407–425. https://doi.org/10.1002/jeab.829