

SOBRE EL PAPEL DE LOS MODELOS EN CIENCIA: UN EJEMPLO

Dr. José E. Burgos¹

RESUMEN

Este artículo ejemplifica la idea de que el papel de los modelos en ciencia consiste no solo en describir y explicar fenómenos, sino también resolver problemas conceptuales. Esta idea proviene de la metodología de las tradiciones de investigación de Larry Laudan, según la cual la ciencia es una actividad de solución de problemas tanto empíricos como conceptuales. Según esta metodología, algunos problemas conceptuales son internos y se refieren a la equivocidad de las categorías básicas de análisis de una teoría. El ejemplo que aquí se examina es la excesiva equivocidad del término "aprendizaje". Tal equivocidad es endémica a todas las teorías del aprendizaje y ha generado otros problemas, como el problema si el condicionamiento operante y el condicionamiento respondiente involucran uno o dos tipos de aprendizaje. Ambos problemas pueden ser resueltos mediante un modelo computacional del condicionamiento que consiste solo de interpretaciones matemáticas de categorías neurales. El modelo resuelve el problema de la equivocidad del término "aprendizaje" eliminándolo, idea que proviene de la metodología de eliminación de términos teóricos de Frank Ramsey. Esto, a su vez, permite una reformulación y solución más satisfactoria (clara, precisa, detallada, parsimoniosa y heurística) del problema de los dos tipos de condicionamiento.

Palabras clave: problemas conceptuales, aprendizaje, condicionamiento operante, condicionamiento respondiente, modelos computacionales, categorías neurales.

1 Universidad de Guadalajara. Correspondencia puede ser enviada a jburos@cucba.udg.mx.

INTRODUCCIÓN ANECDÓTICA

En el curso del mes de enero del año 2009 recibí una invitación del Dr. Nestor Schmajuk (<http://fds.duke.edu/db/aas/pn/faculty/nestor>) a participar en un simposio que estaba organizando en la Universidad de Duke (su adscripción institucional) sobre lo que en ese momento denominó “modelos computacionales de condicionamiento clásico”. El simposio se celebró el 16 y 17 de mayo de ese año, como una sesión de trabajo a puerta cerrada en el aula McDougall de la Universidad de Duke, nombrado así en honor a William McDougall (el funcionalista), quien fundara el primer laboratorio de psicología experimental de dicha universidad (en realidad, era un laboratorio de parapsicología, pero su aproximación al tópico fue estrictamente científica, es decir, naturalista-experimental). Por supuesto, tal invitación fue un gran honor para mí, un reconocimiento de mi trabajo científico por parte de uno de los líderes actuales del área de modelos de redes neurales de condicionamiento (autor del primer libro completamente dedicado al tema). El honor se acrecentó, al punto de éxtasis, cuando me enteré de quiénes eran otros de los invitados, como por ejemplo Allan Wagner (el del modelo Rescorla-Wagner; <http://www.yale.edu/psychology/FacInfo/Wagner.html>) y Ralph Miller (conocido por su modelo comparativo; <http://www2.binghamton.edu/psychology/people/faculty/ralph-miller.html>). Por fortuna, esos estados de éxtasis son efímeros y dan paso a lo más importante: el tema propuesto y cómo podía yo contribuir al mismo. ¿Sobre qué iba a hablarles a esos titanes de la psicología experimental? Mi éxtasis se convirtió en pánico. Por supuesto, tenía claro que Schmajuk me había invitado para que hablara sobre el modelo de redes neurales con el que he estado trabajando por 20 años. Más bien, mi gran duda era *específicamente* qué les iba a decir sobre el modelo que no hubiera sido publicado antes y que por tanto justificara mi presencia en tan importante ocasión.

Inmediatamente pensé en el tema ideal y acepté la invitación proponiendo como título de mi charla el equivalente en inglés de “La distinción operante-respondiente: Una interpretación de redes neurales.” La charla la pensé como una versión modificada del que fuera mi trabajo de ascenso en la UCAB, el cual presenté a finales de los 90. Mi advertencia obvia a Schmajuk fue que mi charla no iba a tratar solo sobre condicionamiento clásico, sino también condicionamiento operante, cómo se distinguían y relacionaban ambos tipos de condicionamiento. Así y como lo mostré en ese trabajo de ascenso, el modelo ofrecía su propio modo peculiar de resolver esta cuestión. La modificación crítica, para efectos de mi presentación, consistió en adoptar una *metodología* explícita, en el sentido filosófico del término, es decir, una forma específica de *justificar* por qué la solución del modelo era *mejor* que otras.

A pesar de mi advertencia, Schmajuk continuó refiriéndose informalmente al simposio como “modelos computacionales de condicionamiento clásico”. Mi pánico se convirtió en frustración y hasta irritación (es interesante cuán rápidamente pueden cambiar las emociones). Ello me forzó a planear el inicio de mi charla así: “En contra del título oficial de este simposio, no hablaré solo de condicionamiento clásico, sino también de condicionamiento operante”. Gracias a mi charla, Schmajuk finalmente cambió el título oficial (para efectos de su publicación por Cambridge University Press) a “Modelos computacionales de condicionamiento”. Solo espero no haberlo irritado por ello. No pude dejar de tener la sensación de que Schmajuk quizás se había arrepentido de haber invitado a un iconoclasta revoltoso como yo. En todo caso, mi frustración e irritación se convirtieron en desacato. En ese momento, me sentí listo para enfrentar el paredón de fusilamiento académico.

La metodología que adopté tiene que ver con lo que dije a continuación en mi charla: “También, a diferencia del resto de las charlas, la mía tratará sobre cuestiones *conceptuales*”. Esta afirmación fue un riesgo calculado, por cuanto yo estaba consciente de que el énfasis principal, sino exclusivo, del área era no solo sobre condicionamiento clásico, sino sobre lo que denominé el papel *descriptivo-explicativo* de los modelos matemáticos en teoría del condicionamiento (y en ciencia en general). Este papel tiene que ver con el uso de un modelo para describir y explicar fenómenos conocidos (otro papel, si se quiere, es predictivo y se refiere a la capacidad de un modelo de predecir nuevos fenómenos, pero lo dejaré de lado en este artículo).

En efecto, el resto de las charlas giró exclusivamente en torno al papel *descriptivo-explicativo*. Sin embargo, aclaré en mi charla, aunque ese papel ciertamente es crucial, no era el único papel de los modelos en ciencia. Otro papel igualmente importante, argüí, era *conceptual o interpretativo*, y se refiere a la formulación y solución de *problemas conceptuales*. Tales problemas son planteados por definiciones de y relaciones entre categorías básicas de análisis de los fenómenos de interés. De esta manera, mi charla desafió uno de los supuestos fundamentales implícitos del área, a saber: el *único* papel científicamente aceptable de los modelos en teoría del condicionamiento es el *descriptivo-explicativo*. De hecho, propuse que cumplir con un papel *conceptual-interpretativo* era un criterio de evaluación de modelos matemáticos. Es decir, la capacidad de formular y resolver problemas conceptuales es una *virtud* que cuenta para la aceptabilidad de un modelo.

Ello, por supuesto, causó desconcierto (quizás irritación) entre los demás participantes, quienes estaban acostumbrados solo al papel *descriptivo-explicativo* de los modelos, y que, por tanto, plantearse cuestiones conceptuales les resultaba algo enteramente foráneo. Por esta razón, sus preguntas resultaron

en algunos casos desatinadas, en otros, ingenuas. Por ejemplo, John Kruschke (<http://www.indiana.edu/~kruschke/>), otro de los participantes, preguntó si el modelo podía simular la inhibición condicionada. Respondí admitiendo que aún no lo sabía, ya que aún no lo había intentado. Más importante aún, mi énfasis no era sobre lo que en ese momento denominé “el registro empírico” del modelo, es decir, la lista de fenómenos que podía simular, sino sobre su utilidad para plantear y resolver un cierto problema conceptual (viz., el problema de los dos tipos de condicionamiento). A pesar de mi aclaración, otros participantes insistieron en preguntar si el modelo podía simular este o aquel fenómeno, ante lo cual me limité a repetir la aclaración casi como disco rayado. Aquellos que entendieron mi énfasis conceptual, como Wagner y Miller, optaron por señalar que mis definiciones eran demasiado simplificadas. Wagner, por ejemplo, comentó que la forma que tenía el modelo de definir una respuesta (ver más adelante) era demasiado simplificada, y que no capturaba la complejidad de las respuestas tal y como eran estudiadas en animales. Mi reacción casi automática fue afirmar que eso podía decirse de *cualquier* modelo (recuerdo vívidamente que Schmajuk expresó su acuerdo, de manera igualmente automática). Ante tal reacción, Wagner calló y no comentó nada más.

El énfasis exclusivo sobre el papel descriptivo-explicativo de los modelos computacionales del condicionamiento, ya para hacerlo más específico, también fue patente en algunos de los comentarios a la versión escrita de mi charla. Uno de los revisores, por ejemplo, hizo cinco comentarios, de los cuales tres se referían a la capacidad del modelo de describir ciertos fenómenos (los tres comentarios empezaban con la frase “Parece que el modelo no describe...”).

UNA METODOLOGÍA: LAS TRADICIONES DE INVESTIGACIÓN

Mi énfasis sobre el papel conceptual de los modelos en ciencia, por supuesto, ya estaba implícito en el antes mencionado trabajo de ascenso. Tal énfasis, debo confesar, ha sido muy cercano a mi trabajo durante toda mi carrera. En ese entonces, sin embargo, ignoraba que tal énfasis era parte central de una metodología que había sido publicada en forma de libro a finales de los setenta, a saber, la metodología de las *tradiciones de investigación*, propuesta por el filósofo de la ciencia Larry Laudan (1977). Al encontrármela un par de años antes de la charla de Duke, vi de inmediato mi énfasis en lo conceptual reflejado (y en esa medida validado) casi a la perfección, y formulado de manera mucho más clara, vívida y detallada, con ejemplos concretos de la historia de la ciencia.

La propuesta principal de Laudan es que la ciencia no es una búsqueda de la verdad, sino una actividad de *solución de problemas*. Esta propuesta, tal y como lo reconoce el mismo Laudan en su libro, no es novedosa. Más bien, lo novedoso es la forma en la cual Laudan *elabora* esa propuesta general, así como

su implicación, por demás controversial, de que la ciencia no es una búsqueda de la verdad. Laudan distingue entre dos tipos de problemas que los científicos tratan de resolver, a saber: problemas empíricos y problemas conceptuales. Los problemas empíricos son preguntas planteadas por nuestra observación de la naturaleza: ¿Por qué llueve? ¿Por qué el cielo es azul? ¿Por qué se asemejan los hijos a sus padres? ¿Por qué los organismos se comportan como se comportan? Estos problemas, plantea Laudan, se resuelven en ciencia mediante *teorías* (aunque no especifica exactamente qué es una teoría para él). En el presente caso, plantear y resolver este tipo de problemas es lo que he llamado el papel descriptivo-explicativo de los modelos matemáticos (supongo que Laudan admitirá que los modelos matemáticos califican como al menos un tipo de teoría).

Por su parte, los problemas conceptuales se refieren a cuestiones de claridad, precisión, coherencia y elaboración de las categorías básicas de análisis dentro y fuera una teoría. Laudan clasifica estos problemas en dos tipos: internos y externos (a una teoría). Los problemas conceptuales internos se refieren a ambigüedades e inconsistencias en las categorías básicas de análisis de una teoría. Un ejemplo muy famoso de la física es el concepto de fuerza en la teoría de Newton, el cual fue duramente criticado por sus contemporáneos por ser demasiado oscuro (la reformulación de la teoría por parte de Joseph Louis Lagrange buscaba, justamente y en parte, clarificar la noción de fuerza; la reformulación de la mecánica Lagrangiana por parte de William Hamilton también buscaba lo mismo, al menos en parte). Los problemas conceptuales externos, por su parte, se refieren a las relaciones lógicas entre dos teorías T1 y T2. Estos problemas se dividen en los siguientes tipos: inconsistencia (T1 implica la negación de T2), incompatibilidad (T1 implica que T2 es poco probable), compatibilidad (T1 no implica nada sobre T2) y apoyo (T1 hace a T2 más probable).

En este trabajo haré énfasis sobre problemas conceptuales internos. Mi propuesta general en este sentido es que *todos* los modelos computacionales del condicionamiento actualmente disponibles (de hecho, todas las teorías del aprendizaje), sufren del mismo problema conceptual interno, y que el modelo aquí defendido resuelve este tipo de problema de una manera drástica. Para elaborar esta propuesta general, primero identificaré el problema en cuestión, para luego mostrar cómo el modelo lo resuelve.

UN PROBLEMA: LA EXCESIVA EQUIVOCIDAD DEL TÉRMINO “APRENDIZAJE”

Todos los modelos del condicionamiento y todas las teorías del aprendizaje actualmente disponibles (excepto, por supuesto, el que presentaré más adelante)

sufrían del mismo problema conceptual interno, a saber: se fundamentan sobre una categoría de análisis excesivamente equívoca. El objetivo de esta sección es ilustrar este problema con ejemplos concretos de la literatura.

La categoría en cuestión es el concepto de aprendizaje. Una revisión casual de la literatura sobre psicología del aprendizaje revela su equívocidad:

“un cambio más o menos permanente en la potencialidad de la conducta que ocurre como resultado de la práctica reforzada” (Kimble, 1961, p. 6);

“el cambio en la conducta o potencial de conducta de un sujeto a una situación dada causado por las experiencias repetidas del sujeto en esa situación, suponiendo que el cambio conductual no puede ser explicado sobre la base de las tendencias de respuesta innatas del sujeto, la maduración o estados temporarios” (Bower y Hilgard, 1981, p. 11);

“un cambio relativamente permanente en el potencial de un organismo para responder que resulta de experiencia previa o práctica” (Gordon, 1989, p. 6);

“un proceso debido a la experiencia que resulta en un cambio relativamente permanente en la conducta que no puede ser explicado por estados temporarios, maduración, o tendencias innatas de respuesta” (Klein, 1996, p. 2);

“una dependencia de la conducta presente del ambiente como una función de una interacción previa entre la actividad sensorial-motora y el ambiente” (Pear, 2001, p. 12);

“un cambio duradero en los mecanismos de la conducta que involucra estímulos específicos y/o respuestas que resulta de la experiencia previa con esos o con estímulos y respuestas similares” (Domjan, 2009, p. 17).

En su revisión del libro de Hilgard y Marquis, Kimble (1961) le dedica todo un capítulo a la definición de aprendizaje, en el cual hace una distinción entre dos tipos de definiciones, a saber: fácticas y teóricas. La primera definición citada arriba, muy popular en psicología, es fáctica. Aunque Kimble la propuso como una definición fáctica más adecuada, es incorrecto suponer que era *partidario* de ella. Más bien, Kimble reconoció que ninguna definición era completamente adecuada, admitiendo también la importancia de las definiciones teóricas. Tales definiciones, según él, buscan “describir mecanismos subyacentes o identificar la verdadera naturaleza del aprendizaje” (p. 8). Entre estas definiciones, identificó definiciones fisiológicas, cognitivas y conductuales (estímulo-respuesta).

Resulta poco discutible, pues, que el término “aprendizaje” es demasiado equívoco. No solo ha sido definido de muchas maneras distintas, sino que algunas de ellas presentan ambigüedades cruciales. La más obvia es que varias

de las definiciones citadas arriba apelan a “potencialidades” o “potenciales” de conducta o de responder de cierta manera. Sin embargo, no queda claro qué son tales potenciales.

Por supuesto, no estoy diciendo que el concepto de aprendizaje deba ser definido de manera *perfectamente* clara y precisa, en términos de condiciones necesarias y suficientes. Ello sería muy difícil (si no imposible) y hasta indeseable. Cierta grado de vaguedad en ciencia es inevitable y hasta deseable, ya que ello permite el cambio conceptual. Solo estoy diciendo que el concepto de aprendizaje es *demasiado* equívoco como para ser útil. Tal grado extremo de ambigüedad es tan inaceptable como la ausencia total de ambigüedad. También estoy diciendo que quizás sea imposible definir el concepto de aprendizaje de una manera lo suficientemente clara y precisa como para que su utilidad sea significativa.

OTRA METODOLOGÍA: ELIMINACIÓN DE TÉRMINOS TEÓRICOS

En lugar de proponer otra definición del concepto de aprendizaje, lo cual solo contribuiría al problema, propongo una solución drástica: eliminar el término “aprendizaje”. Esta solución tiene el mismo espíritu de la metodología propuesta por Frank Ramsey (1929) de resolver el problema de los términos teóricos eliminándolos (aunque su propuesta involucra formalismos que no utilizo aquí). Esta propuesta ha sido retomada en la filosofía contemporánea de la ciencia por los proponentes de la concepción estructuralista de las teorías (e.g., Balzer, Moulines, y Sneed, 1987; Sneed, 1979).

Por supuesto, la eliminación del término “aprendizaje” no se realiza de una manera automática y caprichosa. Más bien, es un *resultado* de la forma en la cual el modelo fue construido, la cual hace *innecesario* el término. Todos los modelos computacionales del condicionamiento clásico actualmente disponibles han sido construidos sobre la base de inferencias hipotéticas de posibles procesos internos, que tienen lugar durante el condicionamiento (e.g., activación de representaciones internas de estímulos, cambios en asociaciones entre representaciones, procesos de atención, consolidación de memoria, procesamiento de información, etc.). El presente modelo, sin embargo, propuesto inicialmente por Donahoe, Burgos y Palmer (1993), fue construido sobre la base de interpretaciones matemáticas de algunos sustratos neurales (tanto anatómicos como fisiológicos) del condicionamiento (e.g., activaciones de unidades neurales, cambios en eficacias sinápticas dependientes del hipocampo y de núcleos dopaminérgicos, etc.).

Una consecuencia importante de ese modo fondo-arriba de construcción es que el modelo resultante no hace uso de ninguna de las categorías cognitivas que se encuentran en el resto de los modelos, incluyendo aprendizaje, entre muchas otras (representación interna, asociación, memoria, atención, motivación,

incentivo, impulso, expectativa, hábito, información, etc.). El modelo como tal solo usa (interpretaciones matemáticas de) categorías neurales (unidad neural, conexión, activación, peso). La suposición central del modelo es que durante el condicionamiento, es decir, durante el cambio conductual observado bajo ciertas contingencias ambientales (relaciones temporales y estadísticas entre estímulos y respuestas), ciertos circuitos neuronales en el sistema nervioso del organismo son afectados de cierta manera. Esta suposición hace innecesario el uso de aquellas categorías cognitivas que se encuentran en otros modelos. Bajo tal suposición, lo único que se requiere para explicar el condicionamiento es modelar cómo esos circuitos cambian como resultado de las contingencias.

Obviamente, el modelo no es menos abstracto que otros, si por “abstracto” entendemos “simplificado”. El modelo aún involucra un alto grado de abstracción así entendida. La diferencia crucial, más bien, reside en las *categorías básicas de análisis* del modelo. Mientras que en otros modelos dichas categorías son de carácter *cognitivo*, en el presente modelo las categorías son de carácter *neural*. Este énfasis sobre categorías neurales hace que las categorías cognitivas usuales se vuelvan *superfluas*. El resultado neto, en efecto, es un modelo neural del condicionamiento, carente por completo de referencias a entidades o procesos cognitivos.

De hecho, el término “aprendizaje” fue el que más se resistió a ser eliminado del modelo. La razón fue que el término solía aparecer en aquella parte matemática del modelo que se denomina “función de aprendizaje”. Esta denominación es típicamente usada en modelos de redes neurales para referirse a la función matemática según la cual se cambian los pesos de las conexiones de una red neural (un peso simula numéricamente la fuerza de una conexión y se refiere a la eficacia con la cual una unidad neural puede activar otra). Hubo, entonces, una fuerte inercia terminológica que me dificultó deshacerme de la categoría. Sin embargo, mientras escribía el capítulo para las memorias del simposio, me percaté de que lo que había hecho con términos tales como “representación interna”, “asociación”, “memoria”, etc., también podía hacerlo con el término “aprendizaje”. En ese capítulo, entonces, ya no hablo de “función de aprendizaje”, sino de “función de cambio de pesos”. La segunda expresión es un tanto más engorrosa que la primera, pero lo que se pierde en eficiencia expresiva se gana en claridad conceptual. Ello, tal y como lo mostré en la sección anterior, debido a que el término “aprendizaje,” al igual que otros que plagan la literatura, no es *técnico*, es decir, se define de múltiples maneras, muchas de ellas obscuras.

El modelo, pues, permite plantear y resolver el problema de los dos tipos de condicionamiento solo en términos de (interpretaciones matemáticas de) categorías neurales, sin apelar a ninguna de las categorías cognitivas que usualmente se encuentran en teoría del condicionamiento. Ello, propongo, soluciona

un viejo problema conceptual interno en teoría del condicionamiento, a saber, la ambigüedad de las categorías cognitivas. Al usar solo categorías neurales, el modelo permite una formulación y solución un poco más clara y precisa del problema de los dos tipos de condicionamiento.

Las categorías neurales del modelo se refieren a estructuras y funciones que constituyen algunos de los sustratos neurales del condicionamiento, tanto operante como respondiente. Por supuesto, esas categorías deben permitir una correspondencia mínimamente clara y precisa con las *categorías conductuales básicas* que constituyen el aspecto empírico de la distinción, en particular, estímulo, respuesta y contingencia (así como las distintas subdivisiones de cada categoría general: estímulo condicionado, estímulo incondicionado, respuesta condicionada, respuesta incondicionada, respuesta emitida, contingencia respondiente, contingencia operante, etc.). Tal correspondencia, sin embargo, se obtiene solo al nivel de la *red neural total* propiamente dicha, no de sus unidades o conexiones particulares.

De esta manera, y a diferencia de otros modelos, en los que se propugna un nivel único de análisis *teórico* no conductual (el nivel de procesos de aprendizaje), el presente modelo propugna *dos* niveles. Por una parte, se propugna el nivel neuro-computacional, el cual se refiere (muy gruesamente) al nivel celular, sináptico y micro-anatómico de sistemas nerviosos naturales, y corresponde al nivel de las unidades neurales y conexiones individuales de una red neural artificial. Por otra parte, se propugna el nivel de *sistema*, compuesto por múltiples realizaciones interactivas del nivel neuro-computacional, es decir, el nivel de la red neural total propiamente dicha, el cual corresponde (también muy gruesamente) al nivel macro-anatómico de sistemas nerviosos naturales.

REFORMULACIÓN Y SOLUCIÓN DEL PROBLEMA

Ya están disponibles todos los elementos mínimos necesarios para aclarar cómo el modelo formula y resuelve el problema de los dos tipos de condicionamiento. Usualmente, el problema teórico fundamental se plantea en términos de la siguiente pregunta: ¿Involucran los dos tipos de condicionamiento dos tipos distintos de aprendizaje o un solo tipo? Según el modelo, ya podrá anticipar el lector atento, esta forma de plantear el problema resulta inadecuada, ya que hace uso de un término no técnico, a saber, “aprendizaje”. Lo mismo puede decirse de otras formas de plantear el problema: ¿Involucran los dos tipos de condicionamiento dos tipos distintos de asociaciones (e.g., ESTÍMULO-ESTÍMULO versus RESPUESTA-CONSECUENCIA) o solo un tipo? Debido a que el modelo prescinde del término “aprendizaje” (y otros, como “asociación”), promueve una reformulación del problema: ¿Tienen los dos tipos de condicionamiento

dos tipos de sustratos neurales o un solo tipo? Veamos ahora cómo el modelo responde esta pregunta, lo cual, por supuesto, depende mucho de a qué se llama en el modelo “sustrato neural”.

En el modelo, de nuevo, los sustratos neurales se refieren a la estructura y funcionamiento de una red neural artificial que pretende simular algún circuito neuronal indefinido de un sistema nervioso natural. Una red neural es un conjunto de unidades neurales computacionales conectadas entre sí de cierta manera. Matemáticamente, al nivel neuro-computacional, una unidad neural cualquiera, llamémosla j , se activa en el momento $t+1$ con un nivel entre 0 y 1, según una función activación que toma como argumento la suma de los productos entre la activación de cada unidad conectada a j y el peso de la conexión correspondiente en t . Por su parte, cada peso es cambiado de t a $t+1$ según una función de cambio de pesos que también depende de las activaciones de las unidades que constituyen la conexión cuyo peso se desea cambiar, así como de los pesos ganados por esa y otras conexiones que coinciden en una misma unidad. En este nivel básico neuro-computacional, el modelo usa *exactamente la misma* función de activación para *todas* las unidades neurales computacionales, y de cambio de pesos para *todas* las conexiones. El modelo, entonces, no hace distinción alguna entre dos tipos de función de activación o de cambio de pesos que correspondan de manera directa a cada tipo de condicionamiento. No hay una función de activación o una función de cambio de pesos para un tipo de condicionamiento, y otra función para el otro tipo.

El modelo, entonces, no hace distinción *teórica* alguna a nivel neuro-computacional entre dos tipos básicos de sustrato neural que subyacen de manera directa a cada tipo de condicionamiento. El modelo postula solo un tipo de sustrato neural que no es ni operante ni respondiente. Si el lector necesita un nombre, lo puede llamar aprendizaje por contingencias. El modelo hace una distinción teórica solo al nivel sistémico de la red total, en forma de distinciones entre dos tipos de relaciones entrada-salida (provocación versus emisión), dos tipos de protocolos de entrenamiento que simulan los dos tipos de contingencias de reforzamiento (estímulo-dependiente versus respuesta-dependiente), y sus respectivos efectos en una red (aumento de la activación de ciertas unidades de salida en presencia de ciertas activaciones de entrada, como resultado de ciertas contingencias).

La Figura 1 muestra la red mínima que en este modelo permite la definición de todas esas categorías conductuales. La llamaré N1 (el modelo permite redes más grandes, pero N1 bastará para el presente análisis). La red consta de dos unidades de entrada o S' (S y S^*), cuyas activaciones simulan estímulos, cuatro unidades ocultas o intermeditarias que simulan sistema interneuronales (S'' , $ca1$, M'' , vta), y dos unidades de salida o M' (R y R^*). La distinción entre dos

tipos de relaciones entrada-salida se basa en relaciones entre las activaciones de entrada y de salida. Las activaciones de entrada no se calculan según la función de activación, sino que son manualmente asignadas como parte de un protocolo de entrenamiento que simula los tipos de contingencias ambientales que se usan en investigación sobre condicionamiento. La activación de una o más unidades S simula efectos neurales (e.g., activaciones de neuronas de la corteza sensorial primaria) de estímulos sensoriales exteroceptivos que se usan típicamente como estímulos condicionados (EC) en condicionamiento respondiente y estímulos discriminativos (E^D) en condicionamiento operante, tales como luces y tonos. Activaciones de S^* simulan la presencia de estímulos biológicamente relevantes típicamente usados como estímulos incondicionados (EI) en condicionamiento respondiente y como reforzadores primarios (E^R) en condicionamiento operante, tales como comida, agua y choques eléctricos.

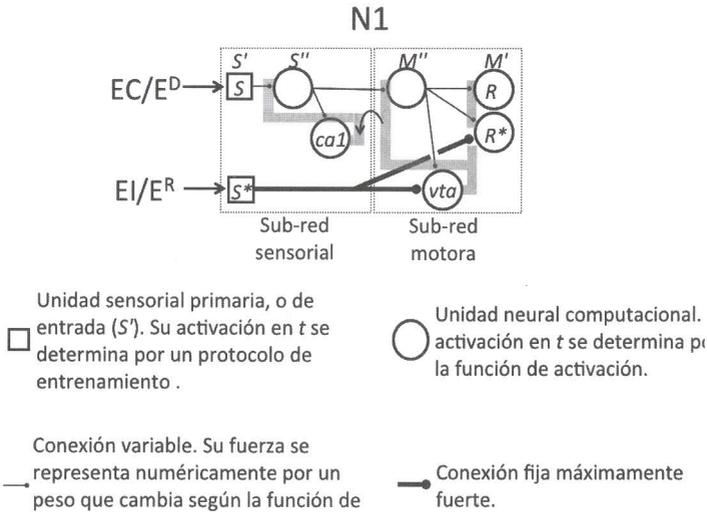


Figura 1. Red mínima para la formulación y solución del problema de los dos tipos de condicionamiento bajo el modelo propuesto por Donahoe, Burgos, y Palmer. EC =estímulo condicionado (condicionamiento respondiente), E^D =estímulo discriminativo (condicionamiento operante), EI =estímulo incondicionado (condicionamiento respondiente), E^R =reforzador primario (condicionamiento operante), S' =capa sensorial primaria, S'' =capa sensorial secundaria, $ca1$ =Cornu Ammonis 1 (el área del hipocampo), M'' =capa motora secundaria, vta =área ventro-tegmental (el núcleo dopaminérgico), M' =capa motora primaria. Áreas sombreadas=señales difusas que afectan cambios en los pesos (la señal proveniente de $ca1$ modula los cambios en los pesos de las conexiones $S-S''$ y $S''-ca1$; la señal proveniente de vta modula los cambios en los pesos de las conexiones $S''-M''$, $M''-vta$, $M''-R$ y $M''-R^*$). Flecha curva=amplificación de la señal de $ca1$ por la señal de vta .

Las activaciones de las unidades M' (R y R^*) representan el responder de N1, y pretenden simular precursores neurales (e.g., activaciones de circuitos neuronales en corteza motora primaria) del funcionamiento de efectores (glándulas y músculos). Digo “responder” y no “respuesta” debido a otra distinción conceptual crucial: activaciones versus respuestas. Una activación es un número real entre 0 y 1, mientras que una respuesta es un evento dicotómico o binario. Por lo tanto, no es correcto identificar activaciones con respuestas. Para poder hablar de respuestas en referencia a N1 es necesario adoptar un *criterio de respuesta* que permita transformar activaciones de salida en eventos binarios. Tal criterio, por supuesto, será en gran medida arbitrario, pero cierto grado de arbitrariedad es inevitable. En otros artículos he usado un criterio según el cual una respuesta es una activación de salida relativamente alta (e.g., mayor o igual a 0.5). De ahora en adelante, hablaré de respuestas en este sentido (el cual fue criticado por Wagner como simplista).

El modelo simula la relación de provocación, característica de los reflejos incondicionados en condicionamiento respondiente, como la activación incondicional de R^* por S^* (i.e., siempre que la activación de S^* sea mayor que 0, se activa R^* con el mismo nivel, independientemente de las activaciones del resto de la unidades, o los pesos de las conexiones variables). Tal activación es posible gracias a que el modelo supone una conexión directa, fija y máximamente fuerte de S^* a R^* (línea negra gruesa de S^* a R^* en la figura). Tal activación será una respuesta si la activación de S^* es mayor que el criterio arriba especificado.

La relación de emisión, característica de las respuestas operantes, resulta un poco más complicada, pero también uno de los aspectos más interesantes de cómo el modelo resuelve el problema de los dos tipos de condicionamiento. *Parte* de esta relación es la *ausencia de activaciones de R en presencia de activaciones de S**. Es decir, la activación de S^* no *causa* la activación de R . Más precisamente, si la activación de S^* es mayor que el criterio de respuesta, R no será activada con ese mismo nivel. Ello debido a que no existe conexión alguna de S^* a R .

La otra parte de la relación de emisión se refiere a la *presencia* de respuestas R (i.e., activaciones de R que satisfagan el criterio de respuesta). Para que R pueda activarse por encima de un criterio de respuesta de 0.5 (o mayor), sin embargo, la red debe ser expuesta a un cierto protocolo de entrenamiento que simule precisamente el tipo de contingencias que promueven el condicionamiento. Tal protocolo debe promover el *incremento de los pesos de las conexiones S-S'', S''-M'', M''-R*. De lo contrario, la activación de R nunca podrá satisfacer un criterio de respuesta de 0.5 (o mayor).

¿Qué procedimiento debería simular ese protocolo? Esta pregunta le plantea un serio dilema al modelo. Por una parte, si fijamos el criterio de respuesta en 0.5, ninguna activación de R satisfará ese criterio si N1 es *ingenua*, es decir,

si los pesos iniciales de todas sus conexiones variables son cercanos a 0 (e.g., 0.01). Ello resultará en activaciones de R igualmente cercanas a 0, por lo que no satisfarán ese criterio y no habrá respuestas R . Por otra parte, si no hay respuestas R , entonces el protocolo no puede simular un procedimiento operante, ya que tal protocolo requeriría de respuestas R (recuérdese que la ocurrencia de una respuesta es indispensable para que se pueda dar una contingencia operante). En consecuencia, N1 no puede ser ingenua, en el sentido de que los pesos iniciales de sus conexiones variables deben ser sustancialmente mayores que 0.

El modelo permite esto de tres maneras. Una manera trivial, por arbitraria, es suponer que los pesos iniciales son lo suficientemente altos como para que las activaciones de R cumplan con el criterio de respuesta. Esta suposición, sin embargo, plantea la pregunta de por qué los pesos iniciales son tan altos. Otra manera, igualmente arbitraria, es suponer que el parámetro de activación espontánea de R es máximo (1.0). Este parámetro es uno de los parámetros libres de la función de activación y determina el nivel de activación de una unidad en ausencia total de activación de entrada. Sin embargo, ello resultaría en una activación de R de 0.3, lo cual no satisfaría el criterio propuesto y redundaría en la ausencia de respuestas R .

La tercera manera es la menos arbitraria, plausible y natural, a saber: pre-entrenar a N1 en un protocolo que simule una contingencia respondiente o, en términos de la nomenclatura usada en la Figura 1, $S-S^*$. En este tipo de contingencia, característico del condicionamiento respondiente, cual la ocurrencia de un EI depende solo de la ocurrencia de un EC. El ejemplo típico es la entrega de varios ensayos de un arreglo anterógrado demorado, donde el inicio del EC precede por un cierto tiempo (el llamado "intervalo entre estímulos") al inicio del EI.

El modelo simula tal arreglo de la siguiente manera. Un EC se define como la supuesta ocurrencia de un cierto cambio físico-químico en el ambiente local de N1 (que no se simula explícitamente) que causa una activación alta de S por un cierto número de momentos temporales (que simula la duración del EC). Un EI se define como la supuesta ocurrencia de otro tipo de cambio físico-químico en el ambiente local de N1 que causa una activación alta de S^* en el último momento del EC. Múltiples ensayos de este arreglo causan un incremento en los pesos de las conexiones variables, lo cual permite que eventualmente una activación alta de S cause no solo una activación alta de R^* (lo cual simula condicionamiento respondiente) sino también una activación alta de R . Ello permitirá que la activación de R alcance el criterio deseado de respuesta. Una vez que esto ocurra, N1 podrá ser entrenada en un protocolo que simule una contingencia operante o, en términos de la nomenclatura usada en la Figura 1, $R-S^*$, donde la activación de S^* dependa de la activación de R .

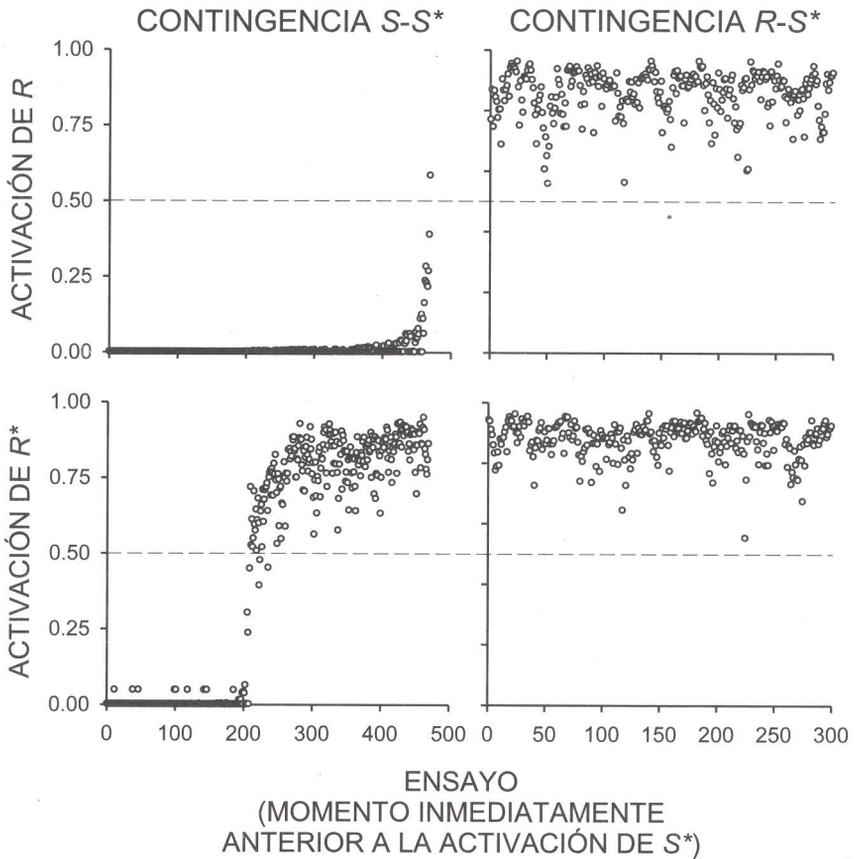


Figura 2. Cambios en las activaciones R (paneles superiores) y R^* (paneles inferiores) como función del número de ensayos de una contingencia $S-S^*$ (respondiente, paneles izquierdos) y una contingencia $R-S^*$ (operante, paneles derechos). Solo se muestran las activaciones del penúltimo momento del EC/ E^D . También se puede apreciar que hacia el final del entrenamiento, la activación de R aumentó lo suficiente como para alcanzar el criterio deseado de respuesta (panel superior izquierdo). En esta fase, la activación de S simula funcionalmente un EC con respecto a la activación de R^* , y un E^D con respecto a la activación de R . En una segunda fase (paneles derechos), se instauró una contingencia $R-S^*$, donde S^* fue activada en el último momento del ensayo si la activación de R era igual o mayor a 0.5. Este protocolo no solo indujo el incremento de la activación de R a niveles cercanos al máximo (panel superior derecho), sino que también mantuvo la activación de R^* a niveles cercanos al máximo (panel inferior derecho).

La Figura 2 muestra los resultados de esa simulación. Todos los pesos iniciales de las conexiones variables de $N1$ fueron fijados en 0.01. Los parámetros libres de las funciones de activación y de cambio de pesos fueron los mismos que en otras simulaciones. En una primera fase, $N1$ recibió ensayos de una contingencia $S-S^*$ hasta que la activación de R alcanzara el criterio de respuesta de 0.5 (línea discontinua). Ello ocurrió luego de 470 ensayos. El resultado de este entrenamiento se muestra en los paneles izquierdos de la figura, donde se puede apreciar que la activación de R^* en presencia de activaciones de S aumentó sustancialmente luego de unos 200 ensayos, a niveles cercanos al máximo, lo cual simula condicionamiento respondiente (panel inferior izquierdo).

COMENTARIOS DE CIERRE

En suma, el modelo mantiene correctamente una distinción *conductual empírica* entre los dos tipos de condicionamiento al nivel de la red total y en referencia a relaciones entre activaciones de entrada y activaciones de salida que simulan los dos tipos de contingencias y sus efectos principales. El modelo, sin embargo, no hace distinción teórica alguna al nivel neuro-computacional de las funciones de activación y cambio de pesos (i.e., no hay dos funciones de activación o de cambio de pesos que correspondan directamente a los dos tipos de condicionamiento). El modelo, entonces, hace una distinción teórica en un sentido (referido al nivel del funcionamiento de la red total), pero no en otro (referido al nivel neuro-computacional). Así, el modelo cumple con un papel conceptual al proponer una mejor solución al problema de los dos tipos de condicionamiento. La solución es mejor porque es más clara y precisa, ya que prescinde de términos problemáticos (e.g., “aprendizaje”, “asociación”, “incentivo”, etc.) y está relacionada de manera más explícita con los sustratos neurales del condicionamiento. También, la solución es más parsimoniosa, ya que apela solo a categorías neurales, en lugar de categorías cognitivas y neurales, tal y como se hace en el resto de los modelos de redes neurales del condicionamiento. En este sentido, el modelo representa un avance respecto a otros en teoría del condicionamiento.

Resulta obvio que la presente propuesta está muy lejos de ser perfecta. Para empezar, las dos metodologías que he combinado son muy controversiales. Ni siquiera está claro si combinarlas, tal y como lo he hecho aquí, resulta técnicamente factible. La eliminación de términos teóricos, en particular, ha sido criticada por ser meramente sintáctica. Hempel (1958), por ejemplo, apuntó que tal eliminación “evita referencia a entidades hipotéticas solo en letra, más que en espíritu” (p. 81). Es decir, la eliminación de términos, una estrategia puramente sintáctica, no garantiza la eliminación del concepto al cual se refiere el término.

Respecto a la forma en la cual el modelo simula el condicionamiento operante, R no puede ser activada sustancialmente de manera espontánea, es decir, en ausencia de *algún* estímulo antecedente. Una activación alta de R requiere de una activación alta de S . Este modelo, entonces, simula condicionamiento operante solo bajo una contingencia operante *discriminada*. El modelo no permite simular condicionamiento operante libre, es decir, sin ensayos discretos explícitamente programados. Ello ciertamente representa una limitación importante del modelo.

Otro comentario es que la simulación implica la necesidad de una contingencia respondiente para que una contingencia operante, y en esa medida el condicionamiento operante, sea posible. Ello no significa que el *condicionamiento* respondiente sea necesario para el condicionamiento operante, si por condicionamiento en estas redes se entiende el cambio en la activación de R^* como resultado de una contingencia respondiente. De acuerdo con esta definición, entonces, la ausencia de unidades R^* implica necesariamente la ausencia de condicionamiento respondiente. Sin embargo, en estas redes tal ausencia no impide simular condicionamiento operante. Si $N1$ tuviera como unidad de salida solo R , aún mostraría los resultados que se muestran en los paneles superiores de la Figura 2.

Además, el modelo predice que el condicionamiento respondiente es sustancialmente más rápido que el condicionamiento operante. Esta predicción es novedosa, pero su confirmación experimental no es posible en la actualidad, ya que aún no existe una forma de hacer comparables las escalas de tiempo involucradas en uno y otro tipo de condicionamiento, ni de ejercer todos los controles experimentales necesarios.

Por último, la contingencia operante fue *simultánea*, es decir, S^* se activó en el mismo momento que R . Esta forma de simular una contingencia operante resulta atípica, ya que por lo general, el reforzador ocurre un tiempo *después* de la respuesta. Aún falta por determinar si el modelo arroja resultados similares con una contingencia operante diacrónica, en la cual S^* sea activada el momento inmediatamente *después* de que R satisface el criterio de respuesta.

REFERENCIAS

- Balzer, W., Moulines, C. U. y Sneed, J. D. (1987). *An architectonic for science: The structuralist approach*. Dordrecht: Reidel.
- Bower, G. H. y Hilgard, E. R. (1981). *Theories of learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Domjan, M. (2009). *The principles of learning and behavior* (6ta. Ed.). Belmont, CA: Wadsworth.
- Donahoe, J. W., Burgos, J. E., y Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, 60, 17-40.
- Gordon, W. C. (1989). *Learning and memory*. Pacific Grove, CA: Brooks/Cole.
- Hempel, C. G. (1958). The theoretician's dilemma: A study in the logic of theory construction. En H. Feigl, M. Scriven y G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem* (pp. 37-98).
- Kimble, G. A. (1961). *Hilgard and Marguis' Conditioning and Learning*. Nueva York: Appleton-Century-Crofts.
- Klein, S. B. (1996). *Learning: Principles and applications* (3ra. Ed.). Nueva York: McGraw-Hill.
- Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Londres: Routledge.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Pearce, J.M., y Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Pear, J. (2001). *The science of learning*. Filadelfia: Psychology Press.
- Ramsey, F. (1929). Theories. En J. Mellor (Ed.), *Frank Ramsey: Philosophical papers*. Cambridge: Cambridge University Press.
- Sneed, J. D. (1979). *The logical structure of mathematical physics* (2da. Ed.). Boston: Reidel.