

Las cuatro causas de la conducta

[Peter R. Killeen](#)

Resumen

La comprensión de un fenómeno implica la identificación de su origen, estructura, sustrato, la función y qué representan estos factores en algún sistema formal. Aristóteles proporcionó una especificación clara de este tipo de explicación, que calificó de causas eficientes (triggers), las causas formales (modelos), las causas materiales (sustratos o mecanismos), y las causas finales (funciones). En este artículo, el marco de Aristóteles se aplica al condicionamiento y el debate cálculo-versus-asociación. La cuestión empírica fundamental es la reducción temprana versus tardía de la información a disposición. La teoría de autómatas proporciona una gramática para modelos de condicionamiento y tratamiento de la información en la que se puede esa restricción representada.

Palabras clave: asociaciones, autómatas, causalidad, explicaciones, modelos

A juzgar si el aprendizaje se explica mejor como un proceso asociativo o de cálculo requiere que se clarifiquen los términos clave. Este ensayo proporciona un marco para la discusión de *explicación*, *asociación* y *cálculo*, sino que deja *el aprendizaje* como un primitivo sin examinar.

Cuatro causas de Aristóteles

[Aristóteles \(trad. 1929\)](#) describe cuatro tipos de explicación. Debido a la mala traducción y la interpretación errónea por "charlatanes eruditos" ([Santayana, 1957](#), p. 238), sus cuatro "because [aitía]" fueron derogadas como un tratamiento incoherente de causalidad ([Hocutt, 1974](#)). Aunque antiguas, las cuatro causas de Aristóteles proporcionan un marco de gran valor para la explicación científica moderna, y en particular para la resolución del debate actual sobre el aprendizaje.

En el marco de Aristóteles, *las causas eficientes* son disparadores, eventos que provocan un "efecto". Este es el significado contemporáneo de la *causa*. Los filósofos como Hume, Mill y Mackie han aclarado los criterios para la identificación de diferentes relaciones causales eficientes (por ejemplo, la necesidad, suficiencia, eventos insuficientes pero necesarios en el contexto de los acontecimientos de otra manera suficientes). Las causas eficientes identifican las primeras partes de una secuencia que son esenciales para las partes posteriores; Nos dicen lo que inicia un cambio de estado ([Jachmann y van den Assem, 1996](#)) "análisis causal etológico" del comportamiento de cortejo de una avispa es un ejemplo de este significado de *causa*.

Las causas materiales son sustratos

Estos son los tipos más comunes de la explicación causal en uso hoy en día, ejemplificadas por la mayoría de las neurociencias y la investigación de imágenes cerebrales. Una vez que la maquinaria ha sido identificada, mucha gente considera que el fenómeno ha sido explicado. Enfoque exclusivo en la maquinaria que se conoce como *reduccionismo*.

Causas formales son modelos

El gran logro de Newton era dar credibilidad a tales modelos con causas materiales ausentes: Para él, no había "corchetes" a la gravedad "hipótesis [relativo a un mecanismo subyacente] no es parte de mi designe", sólo matemáticas al desnudo. Esta fue una situación difícil para Newton de adoptar, para un filósofo mecánico aborrecía oculto (y así ad hoc) cuentas. Newton gustosamente equipó su teoría con los ganchos y los ojos de – causas materiales- pero ninguno pudo idear algo suficiente para mantener a los planetas en sus órbitas.

Las causas formales son mapas lógicos. La forma favorita de Aristóteles fue el silogismo, al igual que en el físico moderno los favoritos son las ecuaciones diferenciales. Tales ecuaciones describen el curso de cambio de un estado a otro, de acuerdo con las condiciones iniciales (causas eficientes), describen la trayectoria completa de cambio.

No importa cuán exitosos son estos modelos formales, no son máquinas: Las ecuaciones matemáticas describen las trayectorias de pelotas de béisbol y los planetas, pero estos organismos no resuelven ecuaciones para proyectar sus movimientos. Los modelos formales de los contribuyentes a esta Sección Especial son mudos sobre la causa eficiente, el sustrato, y la función. Es posible especular sobre los mecanismos subyacentes, y para generar modelos formales de ellos, pero sin datos directos sobre esos mecanismos, los modelos son conjeturas no verificables y por lo general sujetas a cambios como las modas van y vienen, que son ocultas.

Las causas finales son las explicaciones funcionales

"Para reconocer una máquina real, tenemos que tener una idea de lo que se supone que hacen" ([Minsky, 1967](#) , p. 4). Preguntas como "¿para qué sirve?" Y "¿Por qué hace eso?" Llamado a la función (final) causas, la supervivencia del más apto, la teoría de forrajeo óptimo, y las explicaciones intencionales en general dar respuestas pertinentes. La mayor parte de la física moderna se puede escribir en términos de las funciones que optimizan ciertas variables, como la energía. Todas las leyes establecidas en términos de dichas causas se refieren finales óptimos. Los ejemplos más comunes son los rayos de luz por caminos que reduzcan al mínimo los tiempos de tránsito, los animales se comportan de forma que se maximice la representación genética en las generaciones venideras, y los seres humanos se comportan de maneras que maximicen los beneficios para la población. A las causas finales se les dio un mal nombre (*teleología*), ya que eran tratados como errante formal, material, o por causas eficientes. Una razón por la que las jirafas tienen el cuello largo es para hacerles ver follaje alto, pero la causa final no desplaza las explicaciones formales (variación y selección natural) y material (genético), ni es una causa eficiente (Lamarckianismo). Pero ninguna de esas otras explicaciones causales tienen sentido sin la especificación de la causa final. Los biólogos reintrodujeron las causas finales bajo el eufemismo de "mecanismos definitivos", en referencia a las causas eficientes y materiales de una conducta como "mecanismos próximos."

Dos sistemas que comparten las mismas causas finales pueden tener soportes muy disímiles. Los análisis de los análogos evolutivos -tal como las alas de los insectos, aves y murciélagos- proporcionan información útil funcional (relativos, por ejemplo, a las presiones evolutivas convergentes y variedades de estrategias adecuadas para esa función), a pesar de que las alas no son homólogos (es decir, no son evolucionado a partir de un mismo órgano en un antepasado antiguo). Los análisis analógico-funcionales son víctimas de "la falacia analógica" sólo cuando se supone que la similitud de la función implica similitud de eficiencia (historia evolutiva) o causas materiales (fisiológicas). Tales factores de confusión se pueden evitar por la contabilidad para cada tipo de causa por separado.

Las causas eficientes son, pues, las condiciones iniciales de un cambio de estado, las causas finales son las condiciones terminales, las causas formales son modelos de transición entre las condiciones iniciales y terminales, las causas materiales son el sustrato sobre el que actúan estas otras causas.

EXPLICANDO CONDICIONAMIENTO

[Skinner \(1950\)](#) arremetió contra la formal ("teorizar"), material ("neuro-reduccionista") y causas finales ("intencional"), y scientized causas eficientes como "las variables de las cuales la conducta es una función." Él estaba preocupado por causas complementarias que se pueden utilizar en lugar de, en lugar de junto con, su análisis funcional. Pero de todos los fenómenos del comportamiento, el condicionamiento es el menos capaz de ser

comprendido sin referencia a las cuatro causas: La capacidad de estar condicionada ha evolucionado debido a la ventaja que confiere eficientes en el aprovechamiento de las relaciones causales.

Causas Finales

El condicionamiento conforma trayectorias conductuales en los caminos más cortos a refuerzo ([Killeen, 1989](#)). Cuando un estímulo predice un evento biológicamente significativo (un estímulo incondicionado, EE.UU.), los animales mejoran su estado físico por "asociaciones de aprendizaje" entre los acontecimientos externos, y entre esos acontecimientos y las acciones apropiadas. Nichos estables -aquellos habitados por la mayoría de las plantas, animales y hongos- tampoco exigen no apoyar el aprendizaje: tropismos, los impuestos y los reflejos simples se ajustan suficientemente a las regularidades cotidianas de la luz, la marea, y la temporada. Sin embargo, cuando el entorno cambia, es el papel del aprendizaje para volver a colocar la maquinaria para explotar las nuevas contingencias. Mejores explotadores están mejor representadas en la siguiente generación. Este es la causa final -última, en términos de los biólogos- del condicionamiento. Entendiendo el aprendizaje se requiere saber cuáles son las respuestas aprendidas que pueden haber realizado en los ambientes seleccionados por ellos.

Causas eficientes

Estos son los tipos prototípicos de las causas, lo suficientemente importante para la supervivencia de muchos animales que se han desarrollado sensibilidad a los mismos. Los parámetros que son indicadores de las causas eficientes, contigüidad en el espacio y el tiempo, la prioridad temporal, regularidad de asociación, y afectan tanto a similitud de los juicios de causalidad por los seres humanos ([Allan, 1993](#)) como la velocidad del condicionamiento ([Miller & Matute, 1996](#)).

Causas materiales

El sustrato del aprendizaje es el sistema nervioso, lo que proporciona un exceso de riqueza en los mecanismos. Desarrollo de las explicaciones formales y eficientes del condicionamiento puede orientar la búsqueda de mecanismos operativos neuronales. A su vez, la elucidación de esa arquitectura neural puede guiar el modelado formal, como los modelos conexionistas paralelos –redes neuronales- que de emulen diversas funciones cerebrales. Cada una de las cuatro causas es un recurso para la comprensión de los demás.

Las causas formales

Los modelos son subconjuntos propios de todo lo que se puede decir en un lenguaje de modelado. Modelos asociacionistas y computacionales de aprendizaje son formulados en los idiomas de la probabilidad y autómatas, respectivamente. Sus estructuras se esbozan a continuación.

Modelos Asociativos Implicación material, la relación *suficiente* (si C, entonces E, simbolizada como $C \rightarrow E$), proporciona un modelo simplista de ambos causalidad eficiente y condicionamiento. Se sostiene que cada vez que C, entonces también E; falla cuando C y E. Cuando la presencia de una señal (C, el estímulo condicionado o CS) predice con precisión un reforzador (E, de los EE.UU.), la fuerza de la relación $C \rightarrow E$ aumenta. La probabilidad condicional de los EE.UU. dado el CS- $p(E | C)$ -generaliza esta relación todo-o-nada a una probabilidad. Los animales también son sensibles a la presencia de los EE.UU. en la ausencia de la CS, $p(E | \sim C)$; sólo si esta probabilidad es cero es una causa dice que es *necesario* para el efecto. Efectos innecesarios degradan el condicionamiento, al igual que sucesos inesperados crea un observador al cuestionar su comprensión de una situación.

Buenos indicadores de la fuerza de aprendizaje son: a) la diferencia entre estas dos probabilidades condicionales y b) la *diagnostividad* de la CS, $p(E | C) / p(E)$, que es el grado en el que la causa (CS) reduce la incertidumbre respecto a la ocurrencia del efecto (EE.UU.). Como es el caso para todas las probabilidades, la medición de estos condicionales requiere un contexto de definir. Esto puede comprender combinaciones de señales, un entorno físico y la historia de reforzamiento. Refuerzo engendra una actualización de los condicionales, la velocidad del condicionamiento depende del peso de la evidencia implícita que corresponde a los condicionantes anteriores. Las bases de datos de algunos condicionantes, tales como la probabilidad de enfermarse después de experimentar un determinado sabor, a menudo empiezan poco a poco, de modo que una o dos parejas aumentan en gran medida la probabilidad condicional y generan aversiones gustativas. Emparejamientos anteriores del gusto y de la salud, sin embargo, le dará a los condicionantes anteriores más inercia, haciendo que la probabilidad condicional de aumentar más lentamente, y posiblemente proteger al individuo de una aversión al sabor causado por la posterior asociación de la enfermedad con gusto. Estímulos más comunes, tales como formas, pueden ser lentos a condición debido a una historia de exposición que no está asociada con la enfermedad. El Teorema de Bayes proporciona un modelo formal de este proceso de actualización de las probabilidades condicionales. Este es un ejemplo de cómo los subconjuntos de la teoría de probabilidades puede servir como un modelo formal para la teoría de la asociación. Teorías asociativas continúan evolucionando a la luz de los experimentos de manipulación de las variables contextuales, [Hall \(1991\)](#) proporcionan una excelente historia de la restricción progresiva de los modelos asociativos de datos.

Modelos Computacionales Las computadoras son máquinas que se asocian con direcciones de contenidos (es decir, ir a un archivo especificado por una dirección y obtener ya sea un dato o una instrucción). No sólo los equipos asociados, pero las asociaciones de calcular: "Cada máquina de estado finito es equivalente a, y puede ser 'simulada' por, alguna red neuronal" ([Minsky, 1967](#) , p 55.). Las computadoras pueden crear instancias de todos los modelos asociativos de condicionamiento y sus inversas. Para que la metáfora computacional pueda convertirse en un modelo, debe ser restringido a un subconjunto propio de lo que las computadoras pueden hacer, una manera de lograr esto es a través de la teoría de los *autómatas* ([Hopkins y Moss, 1976](#)). La teoría de autómatas es una caracterización formal de arquitecturas computacionales. Una distinción crítica entre autómatas es la memoria: autómatas finitos pueden distinguir solamente los insumos (historias de condicionamiento) que se pueden representar en su memoria interna finita. La representación puede ser ampliada gradualmente añadiendo una memoria externa en forma de empujar hacia abajo las tiendas, los discos regrabables finitas o infinitas cintas. Estas arquitecturas amplificadas corresponden a [\(1959/1963\) de Chomsky](#) gramáticas libres de contexto, gramáticas sensibles al contexto y máquinas universales de Turing, respectivamente. Las máquinas de Turing son modelos de la arquitectura de un ordenador de propósito general que se puede calcular todas las expresiones que son computables en cualquier máquina. La arquitectura de una máquina de Turing es engañosamente simple, dado su poder universal, es el acceso a una memoria potencialmente infinita que la "cinta" le da este poder. Las computadoras personales son, en principio, las máquinas de Turing, instrumentos de silicio cuya universalidad ha desplazado a la mayoría de los instrumentos de metal de una psicología precedente.

La distinción crucial La memoria es también lo que divide a la asociativa de los enfoques computacionales. Disminución precoz de la memoria a disposición de los estados requiere menos memoria que permite la reducción de tarde y más rápido respuestas -reflexivas-, la reducción tarde es más flexible y comportamiento de los animales siempre son representativas de la computación en todos los niveles hasta, pero sin exceder la capacidad de la memoria "inteligente". . La mayoría de los comportamientos humanos son simples reflejos correspondientes a los autómatas finitos. Incluso los repertorios más complejos pueden llegar a ser "automatizado" en la práctica, la reducción de un principio de cálculo intensivo en respuesta a los intentos del niño para atar un zapato a un hábito inconsciente. La adaptación permite la formación llegaría a un precio demasiado alto si no pueden dar lugar a la capacidad de respuesta automático y rápido así. La conciencia de las medidas de adaptación permisos, pérdida del conocimiento permite velocidades.

En teoría asociativa tradicional, la información se reduce a un potencial de acción ("fuerza" de asociación entre el CS y EE.UU.) y almacenados en una base de tiempo real. Tal autómatas finitos con recuerdos limitados son inadecuados como modelos de condicionamiento porque "la *naturaleza* de la representación puede cambiar el tipo de información que contiene puede ser influenciada por [anúnciate diversas operaciones hoc]" ([Hall, 1991](#) , p. 67). Las ratas tienen acceso memorial a más de la historia del medio ambiente y las consecuencias que capturado por bayesiano simple actualización de disposiciones. Miller (por ejemplo, [Blaisdell, Bristol, Gunther y Miller, 1998](#) , véase también el tema) proporcionó un modelo computacional que ejemplifica la reducción de demora en la transmisión.

Si asociativos tradicionales son demasiado simples para ser un modelo viable de condicionamiento, los ordenadores de libre disposición (universales máquinas de Turing) son demasiado inteligentes. Nuestras tiendas de memoria finita caen en algún lugar en el medio. La Teoría de Autómatas ofrece una gramática para los modelos que van desde simples interruptores y los reflejos, a través de complejas asociaciones condicionales, a los sistemas adaptativos que modifican su software a medida que aprenden. El aumento de memoria es a veces esto requiere interna y externa -a veces se encuentra en las marcas, memorandos, y el comportamiento ("gestos facilita la producción de un discurso fluido, al afectar la facilidad o dificultad de recuperación de palabras de la memoria léxica", [Krauss, 1998](#) , p. 58). El contexto es a menudo más de una señal de memoria- que constituye un detallado, de contenido direccionable forma de almacenamiento situada donde es más probable que se necesiten. Tal vez lo más a menudo de lo que pensamos, el medio es la memoria.

La diferencia entre los modelos asociacionista y computacional se reduce a que los autómatas son isomorfos, y esto se correlaciona con la reducción temprana versus tardía de la información a la acción. Ahora, el reto es identificar la clase y la capacidad de los autómatas que son necesarias para describir la capacidad de una especie, y la arquitectura de las asociaciones dentro de los autómatas de manera que basta para describir el comportamiento de los individuos a medida que progresan a través del condicionamiento.

Comprender Explicación

Muchas controversias científicas no se derivan tanto de las diferencias en la comprensión de un fenómeno como explicación de las diferencias en la comprensión: se esperaba un tipo de explicación para hacer el trabajo de otros tipos, y objetar cuando los científicos hacen lo mismo. Centrarse exclusivamente en las causas finales es ridiculizado como teleológico, en causas materiales como reduccionistas, por causas eficientes como mecánico, y sobre las causas formales como "teorización". Pero el respeto a la importancia de cada tipo de explicación, y la correcta posición de las construcciones dentro apropiado dominios empíricos, resuelve muchas controversias. Por ejemplo, las asociaciones son construcciones formales, que no se encuentran en el organismo, pero en nuestras tablas de probabilidad o computadoras, y sólo emular conexiones formadas en el cerebro, y contingencias encontradas en la interfaz de la conducta y el medio ambiente. Las causas finales no son de tiempo invertidos causas eficientes. Sólo un tipo de explicación se hace avanzar cuando se determinan las partes del cerebro que son activas durante el condicionamiento. Provisión de una explicación no reduce la necesidad de los otros tipos. Causas funcionales no son alternativas a las causas eficientes, pero las terminaciones de ellos.

El análisis formal requiere de un lenguaje, y los modelos deben ser un subconjunto propio de ese idioma. El tema de la señal en el análisis formal de condicionamiento, que no es la asociación frente a la computación, sino más bien las circunstancias de reducción información temprana versus tardía, y el papel del contexto, tanto como una señal de recuperación y como la memoria misma. Teoría de Autómatas proporciona un lenguaje que pueda soportar subconjuntos de máquinas adecuadas para modelar estos procesos, desde la simple asociación hasta los repertorios humanos más complejos.

La comprensión es una bestia de cuatro patas, que avanza sólo con el progreso de cada tipo de explicación, y se mueve más gracia cuando esas explicaciones están coordinadas. Es una actividad humana, y es en sí misma susceptible a los análisis cuatripartitas de Aristóteles. En este artículo, me he centrado en el análisis formal de la explicación, y las explicaciones formales de condicionamiento. La comprensión se logrará como tales causas formales se coordinan con el material (los estados del cerebro), eficiente (contextos eficaces) y finales (evolutiva) explicaciones de la conducta.