Notas de Aprendizaje y Comportamiento Adaptable

Arturo Bouzas

2025-03-15

Table of contents

Chapter 1

Notas de Aprendizaje y Comportamiento Adaptable

Prefacio

Este es un libro de notas para los cursos de "Aprendizaje y Comportamiento Adpatable", que se imparten el la facultad de Psicología de la Universidad Nacional Autónoma de México

El proyecto de esta página web fue financiado por el proyecto PAPIME PE302221

Arturo Bouzas.

Chapter 2

Introducción

La forma en que se organizan y presentan..

Chapter 3

Principios de la Selección Natural

> "Nothing in Biology (Psychology) Makes Sense Except in the Light of Evolution". Dobzhansky.

Como vimos en una nota anterior, la influencia de Darwin y el pensamiento seleccionista sobre la Psicología fue enorme. Hoy en día, para entender el comportamiento adaptado y adaptable, resulta indispensable conocer los principios más generales de la teoría de la selección natural. El propósito de esta nota es presentar estos principios. Para un tratamiento extenso del tema, revise las recomendaciones al final de la nota.

Para cualquier observador del mundo natural, dos propiedades le parecerán sorprendentes y merecedoras de contar con una explicación. La primera de ellas es la enorme variabilidad en morfología, fisiología y comportamiento de los organismos que habitan este planeta. La segunda, es que esta variabilidad parece estar finamente ajustada (adaptada) a las características del entorno que habitan los organismos.

3.0.0.1 1. Variabilidad

Un repaso de nuestra experiencia cotidiana, visitas a zoológicos y videos de historia natural en YouTube, nos alerta a la enorme variedad de organismos que pueblan nuestro planeta: desde organismos unicelulares, hongos, medusas, plantas de múltiples tamaños, insectos, peces, aves, mamíferos y desde luego humanos. La variabilidad no es solo a nivel de la morfología, se da también en los mecanismos fisiológicos y en el comportamiento de los seres vivos. Hay organismos que se desplazan, otros que no; organismos que se alimentan de

un solo producto y otros que comen de todo; organismos que regulan su temperatura, otros que no; organismos que se reproducen sexualmente, otros que no; organismos que tienen una sola cría, otros que depositan cientos de huevos. Aún entre una clase, como la de los perros, existen subtipos de todo tamaño, conformación, nivel de actividad, niveles de apego y agresividad.

Actualmente se estiman entre 10 millones a un billón de especies. Solo entre mamíferos existen aproximadamente 5,500 especies, de insectos 91,000 y no deja de sorprender las 250,000 especies de escarabajos (buen tino de los "Beatles"). Solo en el intestino humano hay 140,000 especies virales, y se estiman entre un millón a 1 billón de especies de bacterias en nuestro planeta. Bacterias y virus representan la mayoría de los organismos biológicos.

La variabilidad observada cambia tanto a lo largo del espacio como a lo largo del tiempo. La variabilidad que registramos varía en sí misma dependiendo del lugar del planeta donde se lleve a cabo la observación. Lo que observamos en el desierto de Sonora es muy diferente a lo que observamos en la selva chiapaneca. Fue esta variabilidad lo primero que impactó a Darwin en su recorrido en el Beagle.

A mediados del siglo XIX, también se sabía que la variabilidad cambiaba a lo largo del tiempo. No solo había especies que ya habían desaparecido, sino, y esto era lo importante, ahora se sabía que también habían existido especies de las cuales no se tenían registros pero que habitaron la Tierra en etapas geológicas más tempranas. Era claro que el mundo biológico no había sido "creado" a un mismo tiempo y que podía hablarse de la evolución de los organismos. Gracias a los avances en la datación de capas geológicas, el análisis de fósiles nos permite determinar con mayor precisión la fecha de aparición de las diferentes especies. Estos datos confirman las predicciones de la teoría de Darwin sobre la evolución gradual y la ascendencia común.

Ante el panorama descrito, surgen un número de preguntas, ¿por qué no todos los organismos son iguales?, ¿cuál es el origen de esa gran variabilidad?, ¿cómo dar cuenta de ella y de su distribución en el espacio y en el tiempo? Fueron estas preguntas las que dieron origen a la noción de la evolución.

3.0.0.2 2. Adaptación

Una segunda observación que requiere explicación, es el sorprendente ajuste (adaptación) de las características morfológicas, fisiológicas y de comportamiento , a las características del entorno donde se desenvuelve un organismo. No tan solo existe una inmensa variabilidad, sino que esta está correlacionada con las propiedades de entornos que varían espacial y temporalmente. El ejemplo clásico de una adaptación morfológica lo observó Darwin al visitar los archipiélagos de las islas Galápagos y de Hawaii. Él encontró que la forma del pico de pequeños pájaros genéticamente relacionados, se ajustaba al tipo de alimento en la isla que habitaban. Cómo puede verse en la siguiente figura, el

pico podía ser largo y delgado, apropiado para acceder al néctar dentro de una flor, o corto y fuerte para poder romper y comer semillas.

3.0.1 3. Principios de la Selección Natural

Tres principios se encuentran atrás de la propuesta de Darwin para dar cuenta de la variabilidad y la adaptación de los rasgos de los individuos, en particular de los cambios en la frecuencia relativa de los rasgos en una población : 1. Variabilidad Existe variabilidad en rasgos morfológicos, fisiológicos o conductuales entre miembros de una especie. 2. La variabilidad es heredable 3. Hay una covarianza entre los diferentes rasgos y el número de descendientes dejados por los individuos, la covarianza es parcialmente atribuible al papel causal de los rasgos. Si se satisfacen los tres principios que acabamos de describir, de generación a generación, el rasgo con mayor éxito reproductivo incrementará en frecuencia en la población.

3.0.2 4. Determinantes del éxito reproductivo

3.0.2.1 1. Filtros

Dada cierta variabilidad en un rasgo, la distribución estadística de las características del entorno es la que determina el éxito reproductivo diferencial de dicho rasgo. Una analogía útil para entender el proceso de selección natural es considerar a los entornos como filtros sobre la variabilidad en los rasgos. Son estos filtros los que cambian la frecuencia relativa de un rasgo en una población y determinan su éxito reproductivo diferencial.

Un ejemplo muy sencillo de selección es el juego infantil de inserción de cuerpos geométricos. La cubierta de la caja tiene un conjunto de orificios de diferentes formas: cuadrados, círculos y triángulos. La tarea para el infante es insertar en estos orificios a objetos con forma de cilindros, cubos o pirámides. Para hacer uso de esta analogía de la selección natural, imagine ahora una cubierta enorme de este tipo, con una distribución de orificios de las tres formas, y por otro lado, un saco con una población de los tres objetos con diferentes frecuencias relativas. Esa distribución de objetos sería el equivalente a la "generación número 1" dentro del esquema de la selección natural. Vaciamos el contenido del saco sobre la cubierta y observamos la distribución de objetos que quedaron después de pasar por el filtro de la misma. El número de objetos que sí se ajustaron a un lugar dentro de la cubierta es la distribución de objetos en la segunda generación. Esa distribución depende de dos factores: la distribución original -es decir,el número original de cubos, cilindros y pirámides- y de la distribución de los tres tipos de orificios en la cubierta. Si la cubierta solo tuviese cuadrados y círculos, en la segunda generación solo observaremos cilindros y cubos. Note que no se selecciona el mejor, solo se eliminan los que no pasan a través de los

filtros. Un problema con el ejemplo anterior es que no contempla la creación de nuevos objetos. La selección natural necesita de un proceso que, de generación a generación, produzca nueva variabilidad.

3.0.2.2 2. Determinantes de la variabilidad

La variabilidad puede cambiar intrínsecamente, de generación en generación o puede cambiar por un factor externo que ocurre en una oportunidad. Dos factores están detrás del origen de la variabilidad de generación a generación. El primero son las *mutaciones genéticas aleatorias*, el segundo es la *reproducción sexual*. El primero, las mutaciones genéticas aleatorias, produce modificaciones que pueden tener tres consecuencias: o bien ser letales para el individuo (haciendo que este no pase por ningún filtro evolutivo adicional); o inducir un mayor éxito reproductivo que los demás rasgos heredados (contribuyendo a que el individuo atraviese un nuevo filtro) o bien pueden ser neutrales y no tener ningún efecto sobre el éxito reproductivo. En este último caso, el gen mutado se desliza a lo largo de generaciones y cambia la frecuencia relativa de los diferentes genes: a este proceso se le conoce como *deriva genética* (vea el artículo de xxx para una descripción más detallada). La reproducción sexual es la segunda gran fuerza de variabilidad. La mitad de la conformación genética de cada descendiente proviene del macho y la otra mitad de la hembra; adicionalmente, cada una de las mitades proviene de un muestreo aleatorio de los genes del macho y de la hembra.

La otra fuente de variabilidad es la que se genera cuando un accidente geológico aísla a una población o la divide en subpoblaciones que no pueden interactuar. Es el caso de una erupción volcánica, una separación geológica que forma una isla o algo creado por el humano, como una barda en medio del desierto. El *aislamiento geográfico* produce que un mismo acervo genético separado por una barrera geográfica pueda resultar, mediante la recombinación aleatoria, en dos poblaciones con rasgos diferentes. Dicho proceso inclusive puede derivar eventualmente en la generación de dos especies diferentes que ya no pueden reproducirse entre sí.

3.0.2.3 En resumen,

1. En la teoría de la selección natural los cambios en el éxito reproductivo diferencial se deben a las propiedades (filtros, restricciones) en el entorno de los organismos.

2. La variabilidad aleatoria de los rasgos junto con los cambios en el entorno son el motor de la evolución y de la adaptación de los rasgos de un organismo.

3. Como resultado de la operación de los tres principios de la selección natural, se observa un incremento gradual en el éxito reproductivo de ciertos

rasgos en la población. Este incremento puede describirse como un proceso de ascenso de colina, que resulta en el ajuste de los rasgos a las propiedades estadísticas del entorno.

4. Reservamos el término de *adaptación* a los rasgos resultado del proceso de selección natural. Es importante distinguir entre rasgos que son benéficos para un organismo en el presente de aquellos rasgos cuya existencia es el resultado de un proceso de selección natural.

5. Para entender la evolución de un rasgo, es necesario especificar en detalle los filtros, caracterizados como propiedades estadísticas del entorno. Entre los filtros más generales encontramos:

    a. Limitaciones en recursos
    b. Competencia con organismos de la misma especie y de otras especies
    c. Selección sexual
    d. Selección dependiente de la frecuencia del rasgo

6. Una característica importante del proceso de selección natural es que con frecuencia actúa en entornos que son modificados por el mismo éxito reproductivo de una población. Adaptaciones a un nicho generan un nuevo nicho con un conjunto de nuevos filtros.

7. La teoría de la selección natural debe acompañarse de la especificación de las restricciones, genéticas y físicas, que limitan el tamaño del espacio de los posibles rasgos que son candidatos viables para un proceso de selección. Por ejemplo, físicamente hay una relación posible entre el peso de un animal y el diámetro de sus patas. Es decir, ciertos tamaños de patas en algunas especies no son rasgos viables para ser seleccionados por selección natural debido a las restricciones físicas propias de la anatomía del organismo. En ese sentido, la selección natural no produce la solución perfecta a un problema, sino que resulta en la mejor de las posibles soluciones dado un conjunto de restricciones.

8. Encontramos dos tipos de explicaciones evolutivas:

    a. Explicaciones Causales: tras la observación de un rasgo y de su posible función, estas explicaciones buscan encontrar los cambios en los entornos, las posibles restricciones y la historia de esos rasgos que pueden dar cuenta de su aparición en una población.
    b. Explicaciones de optimización: estas explicaciones están ancladas en las herramientas de la teoría matemática de la optimización y consisten en elaborar modelos del entorno (preferentemente matemáticos) como un problema y derivar su solución óptima dado un conjunto de restricciones. El éxito de estas explicaciones se sustenta en la calidad del modelo de las propiedades estadísticas del entorno que funcionan como filtros y que constituyen el problema a resolver, así como de la

identificación completa de las posibles restricciones de las cuales se derivan las soluciones.

9. Sin embargo, no todos los rasgos observados son el resultado de un proceso de selección natural, hay otros factores que se combinan con el proceso de selección para poder entender la evolución de un rasgo. Estos factores están principalmente asociados con los procesos que resultan en la generación de variabilidad aleatoria.

Chapter 4

Evolución de la Adaptabilidad del Comportamiento: El papel de las restricciones

Si el término "adaptación" se utiliza para designar aquellos rasgos que son resultado del proceso de selección natural, el concepto de "adaptabilidad" del comportamiento se define como la medida en que un comportamiento contribuye al éxito reproductivo en el presente. El primer término hace referencia al origen de un comportamiento, mientras que el segundo hace referencia a la función que cumple un comportamiento actualmente. En particular, la adaptabilidad de un comportamiento se modifica naturalmente en función de cómo este aporta a dos características clave de la vida de los organismos: el metabolismo y la reproducción (Godfrey-Smith, 2017). En cuanto al metabolismo, los sistemas biológicos gastan y agotan la energía almacenada y requieren de un constante abastecimiento de ella. Los procesos evolutivos van moldeando distintas formas exitosas de reabastecimiento a través del éxito reproductivo diferencial del organismo.

El reabastecimiento de energía a través de la acción del organismo está limitado por dos grupos de restricciones, uno del organismo y otro del entorno. La primera condición limitante es que el comportamiento toma un tiempo para llevarse a cabo. Dado que el tiempo disponible es finíto, la suma de todos los comportamientos llevados a cabo es igual al total del tiempo disponible.

$$T = t_1 + t_2 + t_3 \dots t_n$$

Esta es una *restricción lineal* bajo la cual los diferentes comportamientos *compiten* por el tiempo disponible. Este planteamiento asume que no existe tal cosa como una "ausencia de comportamiento", por lo que en toda unidad de tiempo se está generando un tipo de conducta que consume una duración particular de tiempo (aunque este comportamiento sea "reposar" o que el organismo permanezca estático). Bajo este esquema, el incremento de una unidad de tiempo asignado a uno de los comportamientos por parte del organismo, implica una menor unidad de tiempo disponible para el resto de los comportamientos. Cada segundo de tiempo t dedicada al comportamiento 1 es una un segundo menos que puede dedicarse al resto de los n comportamientos.

Una segunda restricción del organismo es su estructura biológica metabólica y neuronal, las cuales limitan el posible rango de comportamientos y uso de recursos energéticos.

El éxito reproductivo de un organismo no solo depende de su habilidad para reabastecerse de energía: es necesario, entre otras cosas, encontrar una pareja con la cual reproducirse, evitar depredadores y escapar del frío y/o calor extremo. Todas estas consecuencias tienen una importancia biológica por su impacto sobre el éxito reproductivo de los organismos y coinciden con lo que comúnmente se conoce como "refuerzos": nosotros seguiremos a Baum y, para enfatizar el origen y el papel en la evolución de estos elementos, les llamaremos *sucesos biológicamente importantes SBI*. En la ecuación 1, corresponde a las acciones ligadas a los diferentes SBI. Cuando hablamos de SBI, no nos limitamos a lo que podría entenderse como necesidades básicas. Selección natural ha operado para filtrar mecanismos que permitan traducir éxito reproductivo diferencial en la detección de todas las variables que pueden estar correlacionadas con él. Por ejemplo, selección natural ha filtrado mecanismos de cognición social que permiten a diversos organismos detectar si una acción particular será bien recibida dentro de un grupo específico de su especie: si bien la aceptación social no se traduce inmediatamente en un mayor acceso a la reproducción o en mayores opciones para nutrir el metabolismo del organismo, a la larga la aceptación social es un SBI dado que facilita las condiciones que permiten al organismo un mayor acceso a ambos elementos. Por lo tanto, para diversas especies, la cognición social es un proxy del éxito reproductivo diferencial, aunque esta variable no satisface directamente una necesidad básica en términos evolutivos.

El segundo grupo de restricciones son aquellas que describen la disponibilidad de recursos de reabastecimiento y su relación a propiedades del entorno de los organismos. A estas restricciones les llamaremos las *propiedades estadísticas del entorno.* Son estás propiedades las que, en conjunción con la restricción lineal sobre el comportamiento, determinan el posible conjunto de distribuciones de acciones que han de ser sometidas al filtro de la selección natural.

Las condiciones del entorno biológicamente importantes pueden ser relativamente constantes en el tiempo o variar a lo largo de él. Dependiendo de esa variabilidad en el entorno, el proceso de selección puede ocurrir en dos escalas temporales diferentes: puede darse entre generaciones o durante la vida individ-

ual de un organismo. Si las condiciones son constantes, el proceso de selección ocurre a lo largo de generaciones y resulta en programas conductuales específicos que en la Psicología se les ha conocido como reflejos, instintos o sesgos y de los cuales se dice que son el resultado de un proceso evolutivo. Podemos decir que el proceso evolutivo "codifica" genéticamente las condiciones constantes del ambiente. Consideren la tarea de diseñar un robot que se mueve en un entorno fijo y que tiene solo una tarea por realizar. En ese caso, la solución del ingeniero consiste en *codificar* las características fijas del entorno en el software del robot. En el curso, vamos a llamar *comportamiento adaptado* al que resulta de un proceso de selección natural: es decir, el comportamiento resultante de un proceso de ajuste lento y gradual a las condiciones del entorno a lo largo de generaciones. En ese sentido, el comportamiento adaptado es un tipo de adaptación.

Sin embargo, los entornos de la mayoría de los organismos y de los robots de servicio y exploración espacial son entornos variables, volátiles e inciertos. Este tipo de entornos requieren de mecanismos que se ajusten a la variabilidad en tiempo real (y no a lo largo de generaciones) para poder ser afrontados con éxito. Por ejemplo, para diseñar un robot (como Wall-E) que pueda funcionar en entornos con SBI variables e inciertos, un ingeniero debe conocer y poder modelar la disponibilidad de fuentes de energía para el reabastecimiento del robot, la distribución de los SBI y la disponibilidad de opciones para que el agente pueda llevar a cabo su meta (que sería recolectar y compactar basura en el caso de Wall-E).

Los entornos variables, volátiles e inciertos se caracterizan por sus propiedades estadísticas. Sin esta variabilidad, no habría selección de mecanismos que permitan la adaptación de los organismos a entornos cambiantes dentro del periodo de sus vidas individuales. La adaptación del comportamiento de los organismos a las propiedades estadísticas del entorno dentro de su lapso de vida individual, la atribuimos a un proceso que llamamos "Aprendizaje".

En resumen,

1. En entornos sin variabilidad, la selección natural puede resultar en comportamientos adaptados a las propiedades invariantes del entorno.

2. En entornos variables e inciertos, selección natural resulta en el surgimiento de mecanismos que permiten "comportamiento adaptable" y que llamaremos *aprendizaje* (¡ojo! este concepto es distinto al de "comportamiento adaptado").

3. Selección natural opera en estructuras, genes y las expresiones de genes (como un sistema nervioso, por ejemplo). Estos son los elementos que permiten y subyacen a los mecanismos del comportamiento adaptable.

En términos evolutivos, la tarea de los agentes es *reducir la incertidumbre* acerca de la ocurrencia de los SBI, *predecir* efectivamente su ocurrencia y llevar

a cabo las *acciones* necesarias para interactuar óptimamente con estos. Así, un primer paso para el estudio del comportamiento inteligente es tener una especificación detallada de las propiedades estadísticas de los entornos de los organismos. Cuatro propiedades estadísticas de los sucesos biológicamente importantes han dominado el estudio del comportamiento adaptable: el tiempo y el lugar de ocurrencia, la relación (correlación, covarianza) con otras características del entorno y la relación con el comportamiento de un organismo. El supuesto más importante que haremos es que si las consecuencias relevantes en el entorno se distribuyen en ciertos tiempos, lugares y están asociadas con ciertas estímulos y comportamientos, un organismo que pueda detectar estas propiedades estadísticas y ajustar su conducta a ellas, podrá asignar más óptimamente su distribución de comportamiento a las metas en competencia.

Los SBI ocurren de forma incierta y la tarea para los organismos es descubrir si su ocurrencia está ligada a ciertos tiempos o a ciertos lugares. Pueden también formar parte de una estructura causal, y en esos casos, la tarea de los organismos es descubrir con qué propiedades del entorno están vinculados. Adicionalmente, los SBI pueden depender de las acciones de los agentes, en cuyo caso, la tarea es descubrir cuál acción es responsable de su ocurrencia.

Una propiedad adicional de los entornos es que las cuatro posibles regularidades que acabamos de detallar (características estadísticas del entorno), pueden variar a su vez de acuerdo a otros estados externos del mundo de los agentes (noche y día, escuela y casa). En los libros de texto, a la adaptación de los organismos ante estas propiedades del contexto se les estudia bajo el nombre de control de estímulos.

Un problema de adaptación más difícil es cuando las propiedades estadísticas del entorno cambian sin señales externas, como resultado de estados ocultos al agente. En estos casos, la tarea es determinar si el cambio observado en la propiedad estadística del entorno constituye un cambio aleatorio o si realmente se trata de una modificación en el estado particular del mundo. En la literatura clásica, a este problema se le ha estudiado bajo el nombre del fenómeno de la extinción.

Hasta el momento hemos supuesto, al igual que la gran mayoría de los teóricos hasta los años 60s del siglo pasado, que los organismos solo detectan sucesos individuales. Revisaremos evidencia de que en adición, los organismos detectan y se adaptan a una característica de segundo orden de los entornos: la tasa de ocurrencia de sucesos individuales en su entorno, la cual se define como el número de sucesos por unidad de tiempo. Por ejemplo, las abejas seleccionan su tiempo de estancia en diferentes lugares de un jardín en función de la tasa de encuentro con flores con néctar. Veremos que la noción de tasa de ocurrencia juega un papel muy importante en las explicaciones contemporáneas del comportamiento.

Finalmente, podemos considerar una propiedad de los entornos que puede considerarse de tercer nivel. Nos referimos a la incertidumbre acerca del tiempo,

lugar, covarianzas y tasas de los SBI. Los camiones de transporte público no pasan siempre a la misma hora (incertidumbre de tiempo), en el mercado no siempre se encuentra la misma fruta (incertidumbre de lugar), no llueve siempre que está nublado (incertidumbre de covarianzas) y un jugador de fútbol no alcanza siempre la misma tasa de goleo en una temporada (incertidumbre de tasas). Hay una relación que podemos describir como una probabilidad condicional en el sentido de que la probabilidad de un evento en sí misma depende (o se encuentra condicionada) a otra distribución de probabilidad. Esta descripción nos sirve para distinguir entre cuando la incertidumbre de un evento es esperada o inesperada. La incertidumbre es esperada si existe una única distribución de probabilidad que la describa. Es el caso cuando se lanza una moneda sin imperfecciones al aire, donde sabemos que la probabilidad de que caiga águila o sol es de 0.5. Por otro lado, la incertidumbre es inesperada cuando los parámetros de la distribución cambian de acuerdo a una segunda distribución de probabilidad: en este caso se habla de una probabilidad condicional. Un caso sería una moneda cuya probabilidad de que caiga águila cambia a lo largo del tiempo de acuerdo a otra probabilidad, como la distribución de la velocidad del viento en un entorno, por ejemplo.

En conclusión, si asumimos que la teoría de la selección natural es una descripción correcta y si los organismos y el entorno operan bajo las restricciones descritas, podemos concluir que el objeto de estudio natural de la Psicología es el estudio de la adaptabilidad del comportamiento. En otras palabras, la Psicología estudia distintos comportamientos y los explica en función de su capacidad para brindarle acceso al organismo a sucesos que le son biológicamente significativos.

En el curso veremos cuales son las soluciones óptimas a los diferentes problemas de adaptación generales que encaran los organismos, así como los distintos mecanismos que posibilitan alcanzar tales soluciones a distintas especies. Veremos que para lograr el objetivo de entender el comportamiento adaptable es fructífero iniciar con un detallado análisis del problema de adaptación en cuestión. Esto último requiere modelar tanto las propiedades estadísticas del problema como las restricciones que imponen los posibles mecanismos de las distintas especies sobre la solución óptima.

Chapter 5

Asignación de Crédito

Al interactuar con su entorno, un agente se encuentra con un constante flujo de estímulos y respuestas que se despliegan en el tiempo. Algunos de los sucesos que encuentra son biológicamente significativos, importantes para su éxito reproductivo. El encuentro inesperado con un suceso biológicamente importante echa a andar dos mecanismos: Uno que controla la respuesta inmediata al SBI y un segundo mecanismo que permite predecir su futura ocurrencia. Considere un organismo que encuentra un inesperado pedazo de comida o un depredador. El primer mecanismo le permite al organismo manipular y consumir la comida, o huir y escapar del depredador. El segundo mecanismo, el que posibilita predecir y controlar un SBI, implica la existencia de una estructura causal en el entorno del organismo: esto es, que existen sucesos que predicen o respuestas que producen la comida o evitan al depredador. La tarea para el agente es seleccionar, dentro de un número gigantesco de posibilidades, a cuál suceso o respuesta atribuirle la ocurrencia de un SBI. A este problema de adaptación se le conoce como el de la *asignación de crédito*.

El vasto espacio de posibles candidatos para la asignación de crédito de un SBI incluye la hora a la que ocurre, dónde ocurre, el enorme grupo de sucesos que lo acompañan o los comportamientos que un organismo genera; pero también podemos incluir momentos, espacios, sucesos y comportamientos que ocurrieron en cualquier momento previo. La comida que un perro callejero se encuentra en una banqueta puede deberse a un transeúnte que el perro percibe en ese momento alejándose de la comida, o a miles de posibles transeúntes que la tiraron en un tiempo cada vez más alejado de su encuentro con el alimento, pero pudo deberse a alguien que la tiró desde un transporte público un segundo antes, o diez minutos antes o un día antes.

Para hacer más manejable la asignación de crédito ante las limitaciones de nuestras observaciones y la riqueza de candidatos, selección natural filtró mecanismos que llamamos *"sesgos inductivos",* los cuales logran dos cometidos: primero, reducen el espacio de candidatos a asignación de crédito y, segundo, establecen

un orden de evaluación para poner a prueba a los candidatos del espacio más reducido en un momento posterior. Los sesgos pueden ser el resultado de la codificación genética de propiedades del entorno bajo el cual evolucionó la especie del organismo o el resultado de su experiencia y aprendizaje individuales.

Históricamente, la contigüidad entre sucesos fue el primer sesgo en recibir atención. El sesgo consiste en suponer que la "contigüidad" entre un estímulo o una respuesta y un SBI es una regla evolutiva muy útil para reducir el espacio de opciones de asignación de crédito. El espacio de asignación de crédito se reduce a sólo aquellos eventos contiguos con el SBI. Si al momento que el perro callejero encontró la comida, este prestaba atención a una ambulancia que pasaba con la sirena encendida y a un transeúnte vestido como estudiante universitario, su espacio de asignación de crédito se reduciría a esos dos sucesos. Para seleccionar entre ellos dos, operaría un segundo sesgo que veremos en una sección subsecuente.

Al inicio del siglo XX, Pavlov le dio sentido experimental y conceptual al estudio de este sesgo. El propósito de los experimentos de Pavlov fue establecer la importancia de la contigüidad en la formación de nuevas asociaciones entre estímulos previamente neutrales y respuestas reflejas. El protocolo, representado en la Figura x, consistió en presentarle a un perro un estímulo auditivo seguido por acceso a comida. Pavlov midió la salivación ante la comida y ante el estímulo auditivo, antes y después de haber sido presentado junto con la comida. Encontró que aparear el sonido a la comida, resultó en que el perro salivaba ahora no tan solo a la comida, sino también al sonido. Al sonido se le conoce como *estímulo condicionado EC* y a la comida como *estímulo incondicionado EI*.

En los primeros protocolos experimentales se consideraba solo un candidato al cual asignar crédito (como un tono) y la manipulación experimental era una imprecisa medida de contigüidad que implicaba diferentes relaciones temporales entre el EC y el EI. Los siguientes son los protocolos más empleados: Ver Figura.

En estos protocolos se encontró que la medida de condicionamiento disminuye conforme incrementa el tiempo entre la terminación del EC y el inicio del EI. A esta relación se le llamó el *gradiente de la demora*. Dependiendo de la preparación, después de menos de un minuto de intervalo entre EC y EI no se observaba aprendizaje. Adicionalmente, si el EI se presentaba antes del EC (procedimiento huella) no se observaba aprendizaje. Más adelante veremos que la historia es más compleja que este resumen, pero por el momento es suficiente que se tenga claridad sobre estos resultados.

5.0.1 ¿Es la contigüidad una condición necesaria para el aprendizaje?

Nos podemos preguntar si la contigüidad es el único sesgo que reduce el espacio de candidatos a la asignación de crédito. Para darle respuesta a esta pregunta

se utilizan dos estrategias: la primera consiste de protocolos experimentales en los que dos o más estímulos igualmente contiguos con el SBI compiten por la asignación de crédito. Este protocolo nos permitiría demostrar si la contigüidad es un factor **suficiente** para reducir el espacio de candidatos en asignación de crédito: en el sentido de que, si ambos estímulos son igualmente contiguos, pero el organismo sólo aprende sobre uno de ellos, esto demostraría que la contigüidad no es una condición suficiente para el aprendizaje. La segunda estrategia se trata de protocolos en los cuales se modifica la demora de la presentación del SBI para observar si la asignación de crédito se mantiene. Esta clase de protocolo nos permitiría darle respuesta a la pregunta de si la contigüidad es una condición **necesaria** para el aprendizaje. John García condujo justo estos experimentos. Inicialmente, a partir de una observación accidental trabajando con los efectos de radiación sobre ratas, García encontró que las ratas dejaban de comer y generaban una aversión a su dieta habitual a pesar de que el efecto de la radiación se presentaba mucho tiempo después de la ingesta de la comida.

La figura x muestra el protocolo del experimento de García. A todos los animales se les daba acceso a un bebedero con agua azucarada en el cual cada contacto detonaba la presentación de un tono. De esa forma había un compuesto conformado por dos estímulos: un tono (EC) y el agua dulce (EI). A la mitad de los animales se les daba una descarga eléctrica con cada lengüetazo que daban al bebedero, mientras que a la otra mitad de los animales se les inyectaba una sustancia que producía un malestar estomacal. García encontró que las ratas que recibieron las descargas eléctricas no dejaron de beber el agua azucarada, pero sí evitaban tocar el bebedero cuando este producía el tono; mientras tanto, las ratas con malestar estomacal dejaban de beber el agua dulce, pero no presentaban aversión al tono. Este experimento muestra que la naturaleza del SBI determina los elementos que entran en el espacio de asignación de crédito. Para las ratas, igual que para otras especies omnívoras, como la nuestra, cuando el SBI es un malestar estomacal, el espacio de elección está conformado por elementos con sabor, pero no por elementos visuales o auditivos. Al sentirnos mal del estómago, lo primero que hacemos es buscar qué comimos, aunque nuestra última comida haya sido muchas horas antes. A este sesgo se le conoce como el sesgo de *relevancia biológica*.

Las ratas aprenden a evitar el sabor asociado con enfermedad aun cuando existen largas demoras (horas) entre la experiencia del sabor y la presentación de la enfermedad. Sin embargo, el que la *contigüidad no sea necesaria*, no significa que no sea un factor. En subsecuentes experimentos que manipularon la duración entre el consumo del alimento y la enfermedad, se encontró también un gradiente de demora en el cual la aversión aprendida al sabor incrementa en función de la reducción de los intervalos entre la presentación del alimento y el EI. Una evidencia adicional sobre el papel de la contigüidad la encontramos en estudios que presentan al organismo dos sabores antes de que este atraviese su experiencia de enfermedad. En estos estudios se ha encontrado que la aversión se genera al sabor que es temporalmente más cercano a la sensación de malestar.

Usando la misma preparación de aversión a sabores de García, se encontró otro sesgo importante que determina cuál de los elementos en el espacio de candidatos a la asignación de crédito es considerado primero. En experimentos en los que se presentan dos sabores, uno novedoso y otro familiar, ambos igualmente contiguos con la enfermedad, las ratas aprenden a evitar solo el sabor que era novedoso. A este sesgo se le conoce como el sesgo de la *novedad*.

El sesgo de la relevancia biológica es evolutivo. Para especies como la rata que son omnívoras y viven principalmente en la oscuridad es importante detectar qué alimento es tóxico por su sabor. Otras especies como las palomas, que habitan nichos ecológicos diferentes, no generan aversión a los sabores. Para estas especies, la dimensión relevante es la estimulación visual y no el sabor del alimento. La coevolución entre aves y polillas ejemplifica la importancia de la relevancia biológica. Las polillas son un alimento para ciertas aves; por otro lado, la selección natural resultó en algunas especies de polillas que son tóxicas para las aves. Esta toxicidad es identificable a través de señales visualmente perceptibles, gracias a lo cual, las aves pueden desplegar su sesgo de relevancia biológica hacia los estímulos visuales y aprender a evitar este tipo de polilla. Simultáneamente, otro grupo de polillas no tóxicas evolucionaron para tomar ventaja de ese mismo sesgo de las palomas y desarrollaron patrones visuales similares a los de las especies tóxicas para evitar ser depredadas. Poner figura.

En resumen, los estudios de aversión a sabores sugieren que: 1. Contigüidad no es una condición necesaria para el aprendizaje. 2. Sin embargo existe un gradiente temporal y hay una mayor aversión al sabor más cercano al malestar estomacal. 3. Existen sesgos biológicos que generan una predisposición a considerar sólo ciertos estímulos para asignación de crédito, los cuales dependen del suceso biológicamente importante, como por ejemplo, sabor para enfermedad estomacal en omnívoros y estímulos visuales para aves. 4. Un importante sesgo adicional es priorizar sucesos que son novedosos (o sorprendentes) dentro del proceso de asignación de crédito. 5. La contigüidad es uno de los sesgos, pero no constituye una condición necesaria para el aprendizaje.

5.0.2 ¿Es la contigüidad una condición suficiente para el aprendizaje?

A finales de los años 60s del siglo pasado, un grupo de investigadores, entre los que destacan Leon Kamin, Robert Rescorla y Allan Wagner, condujeron un grupo de experimentos dirigidos a darle respuesta a la pregunta sobre si la contigüidad es una condición suficiente para el aprendizaje. En estos experimentos se presentó un compuesto de dos o más estímulos (condicionados), igualmente contiguos con el suceso biológicamente importante, el llamado estímulo incondicionado (EI). Un ejemplo de un compuesto de estímulos es la presentación simultánea de una luz y un tono, o la combinación de una figura visual y un color.

5.0.2.1 Ensombrecimiento

Los sucesos que anteceden a un suceso biológicamente importante regularmente están compuestos de estímulos que varían en diferentes dimensiones. Un perro que los amenaza, no solo ladra y gruñe, tiene también cierto color, ciertos ojos y cierta boca. Si les llegara a morder, todas estas características del perro estarían contiguas con el suceso aversivo de la mordida. Si la contigüidad fuese suficiente para el aprendizaje, todas y cada una de las características del perro se convertirían en predictores de un ataque. Reynolds puso a prueba esta conjetura con un sencillo experimento. A dos palomas se les entrenó a discriminar entre dos teclas a las que podían picar. Una de las teclas generaba acceso a un comedero, la otra no. Las teclas estaban iluminadas por un compuesto de dos estímulos que variaban en color o forma. La tecla positiva era un triángulo blanco sobre un fondo rojo, la tecla negativa era un círculo blanco en un fondo verde. (Ver figura). Después de que los animales habían aprendido a responder solo a la tecla positiva, se le presentaron los cuatro estímulos por separado. Se encontró que las palomas responden solo a uno de los dos estímulos del compuesto positivo. Una paloma respondía a la figura, la otra al color.

La importancia del experimento radica no solo en la demostración de que la contigüidad no es una condición suficiente, sino en la ilustración de un principio que será clave en el curso: la *competencia* entre elementos, sean estímulos o respuestas. El experimento de Reynolds ilustra que los estímulos presentados en forma simultánea dentro de un compuesto compiten entre ellos por la asignación de crédito del organismo. En ese sentido, la asignación del crédito a uno de los estímulos por parte del organismo implica la no asignación de crédito al otro estímulo presente. Retomando nuestro ejemplo, si el perro los ataca, para algunos de ustedes el predictor del ataque será el gruñido, para otros será el color y para otros será la raza. Cuando es la primera experiencia con el compuesto de estimulación, los factores que determinan cuál elemento gana incluyen la sobresaliencia de los estímulos y su novedad. La siguiente pregunta es si la historia del organismo con uno de los elementos del compuesto afecta la asignación de crédito. A continuación, veremos una serie de experimentos que sugieren que una vez que se asignó el crédito a un elemento, los organismos dejan de considerar a otros elementos como candidatos.

5.0.2.2 Bloqueo

Imaginen que, después de un par de experiencias visitando restaurantes, ustedes aprenden que un mantel de tela es un buen predictor de la calidad de la comida de un lugar. En su visita a un nuevo restaurante, las mesas de este tienen manteles de tela, pero adicionalmente el restaurante tiene música clásica de fondo. La calidad de la comida es igualmente buena a la del último restaurante con manteles de tela que visitaron, pero en este caso la comida fue contigua tanto con el mantel de tela como con música clásica. ¿Habrán aprendido que la música clásica es un predictor de la buena comida? Para darle respuesta a

esta pregunta, tendrían que observar si al verse forzados a escoger entre dos restaurantes sin manteles de tela, seleccionarán aquel que tiene música clásica sobre el que no la tiene. Veremos que los experimentos indican que una vez que se asignó el crédito de un SBI a un elemento de un compuesto, los otros elementos del compuesto no adquieren ningún crédito.

En 1969 Kamin corrió el primer experimento evaluando la intuición anterior. A dos grupos de ratas se les presentó un compuesto de luz y tono seguido de una descarga eléctrica. Ver Figura. Para el grupo experimental, en una fase anterior se le presentaba la luz seguida de la descarga eléctrica. En la tercera fase, de prueba, se le presentaba el tono sin la luz para evaluar qué tanto habían aprendido las ratas acerca de él. Noten que para los dos grupos, el tono antecede a la descarga eléctrica. La única diferencia entre los dos grupos fue la experiencia previa de la luz con la descarga eléctrica. Kamin encontró que a pesar de que para los dos grupos el tono aparecía contiguo con la descarga eléctrica, las ratas del grupo con el entrenamiento luz - descarga eléctrica no mostraron evidencia de que el tono recibiera ningún crédito por la presentación de la descarga eléctrica. Se dice que la experiencia con la luz bloquea el aprendizaje acerca del tono. De la misma forma, en nuestro ejemplo previo, el mantel de tela bloqueaba el aprendizaje acerca de la música clásica. Estos experimentos muestran que el grado de aprendizaje acerca del elemento de un compuesto seguido de un SBI, depende del grado de aprendizaje adquirido previamente por el otro elemento del compuesto. Una forma de interpretar estos resultados es que los elementos compiten por la asignación de crédito en función de si uno de ellos ya es un predictor del suceso biológicamente importante. El fenómeno de bloqueo es evidencia adicional de que la contigüidad entre un estímulo y un refuerzo no es una condición suficiente para el aprendizaje.

Chapter 6

Asignación de Crédito para Respuestas

El acceso a sucesos biológicamente importantes es fundamental para la supervivencia y reproducción de los organismos. Aquellos organismos que puedan predecir confiablemente la ocurrencia de los SBI tienen una ventaja comparativa en términos de su éxito reproductivo. Aprender que un cielo encapotado predice una fuerte lluvia le permite a un individuo anticiparse y prepararse correctamente para ella. De igual forma, escuchar un rugido le permite a una presa prepararse para el caso de un posible ataque. Sin embargo, el individuo no tiene control sobre lo nublado del cielo, ni sobre la presencia del depredador dado el rugido. Puede predecir cuándo lloverá, pero no puede alterar el que llueva; puede predecir que detrás del rugido esté un depredador, pero no puede modificar su presencia.

Uno de los saltos importantes en la historia evolutiva fue la emergencia de mecanismos biológicos que, a través de la acción y la interacción con el entorno, permiten a los organismos *controlar* la ocurrencia de sucesos biológicamente importantes. Estos mecanismos se encuentran estrechamente asociados a un componente específico de la estructura causal de los entornos: las relaciones que describen cuáles acciones de un organismo son exitosas para obtener mayores opciones de acceso a SBI. Un ejemplo en nuestra especie de estas **"elaciones que describen las acciones exitosas para acceder a mayores opciones de SBI"** son los contratos laborales: en estos se contienen las reglas que especifican las acciones a seguir para acceder a un monto de dinero (lo que equivale a mayores opciones de SBI para nuestra especie). Otros ejemplos son: las reglas que especifican qué acciones llevar a cabo si se desea tomar un transporte público; las reglas que definen las acciones requeridas para iniciar una relación amorosa; las reglas que especifican a su mascota qué acciones le otorgan una comida especial; las reglas que especifican a cada especie las acciones que facilitan su acceso a alimentos, así como los actos que les permiten escapar y evitar

a sus depredadores.

Desde la psicología, nos preguntamos cómo un organismo logra reconocer dichas estructuras causales: específicamente, cómo puede determinar qué acción específica, entre muchas posibilidades, es la responsable del resultado deseado (el SBI). En los libros de texto, al estudio de la respuesta a esta pregunta se le conoce como *condicionamiento instrumental o condicionamiento operante*. En estas notas abordaremos su estudio con base en el mismo grupo de principios con los que abordamos los resultados de los protocolos de condicionamiento clásico.

Antes de describir cómo se aplican los mismos principios a los fenómenos de condicionamiento instrumental, es conveniente revisar el estudio original del que surgió esta área de investigación: siguiendo el mismo proceder que seguimos para entender el condicionamiento clásico. Al inicio del siglo XX, Edward Thorndike condujo una serie de estudios con gatos. Él diseñó una variedad de cajas experimentales, de las que un gato encerrado podría escapar activando dispositivos como un cerrojo o una palanca. (Ver figura). La medida del aprendizaje era el tiempo que le tomaba al gato para escapar de la caja. De los datos mostrados en la figura x, puede verse que el tiempo que le tomaba escapar al gato disminuyó conforme aumentaba el número de ensayos en los que se le encerraba. Al inicio, los gatos intentaban un número grande de respuestas hasta que accidentalmente operaban el dispositivo que abría la puerta. Después de algunos ensayos, el gato empezaba a activar el dispositivo de escape inmediatamente después de que se le metía a la caja. Thorndike caracterizó esta ejecución como una de *ensayo y error*. El gato intentaba diferentes respuestas (ensayos) y las descartaba si no lo llevaban a salir de la caja (error).

Es posible identificar que el resultado de los experimentos de Thorndike está compuesto de dos observaciones. La primera es el conjunto de respuestas que lleva a cabo el gato antes de emitir la respuesta correcta. La segunda observación es que después de varios ensayos, el gato ejecuta de forma casi exclusiva e inmediata la respuesta que fue exitosa para escapar de la caja. Para entender estas dos observaciones, recordemos que en el capítulo anterior vimos que el encuentro inesperado con un suceso biológicamente importante (SBI) echa a andar dos mecanismos: uno que controla el comportamiento apropiado para la interacción con y búsqueda adicional del SBI, y un segundo, que permite predecir y controlar su futura ocurrencia.

Para analizar los dos principios que ilustra el comportamiento de ensayo y error, recordemos que los sesgos inductivos pueden dividirse en dos clases:

1. Aquellos que determinan qué elementos -en nuestro caso respuestas- conforman el espacio de candidatos a la asignación de crédito.
2. Los sesgos que determinan cuál elemento dentro del espacio se debe considerar primero.

En el caso del primer sesgo, el que delimita el espacio de respuestas candidato a la asignación de crédito, las respuestas inducidas por el SBI juegan un papel

equivalente al de las mutaciones y la recombinación genética dentro del proceso de generación de variabilidad en la teoría de la evolución. Las respuestas del organismo y la variabilidad genética coinciden en que ambas generan el espacio de opciones seleccionables (candidatos) dentro de los procesos de selección de los que forman parte. En la teoría de evolución, un conjunto de genes creado por las mutaciones y la recombinación genética es sometido a un proceso de selección por los cambios en el entorno; en la la teoría de los sesgos inductivos, un conjunto de respuestas generadas por un organismo en su interacción con el entorno es sometido a un proceso de selección por el sesgo inductivo del organismo. Por otra parte, el segundo sesgo referido, aquel que establece el orden de prioridad para evaluar las respuestas candidato, es equivalente a los procesos específicos de selección natural. De la misma forma en la que la selección natural se dan procesos bien definidos para descartar y conservar genes particulares de entre un amplio espacio de candidatos, existen procesos bien definidos a nivel de los sesgos inductivos (de la segunda clase) del organismo, que describen cómo este prioriza, descarta y conserva las respuestas de entre su espacio de candidatos.

Para entender los resultados de sus experimentos, Thorndike propuso un principio que se conoce como *La ley del Efecto*, la cual establece que: *"En la presencia de un estímulo (situación, contexto) pueden ocurrir una multitud de respuestas. Aquella que vaya seguida de un estado de cosas satisfactorio tendrá que ser la que se asocia (conecta, selecciona) con el estímulo."*

La ley del efecto de Thorndike prioriza a la contigüidad como el factor que determina tanto el espacio de candidatos, como el orden que establece cuáles elementos evaluar primero. La ley no toma en cuenta el origen de las respuestas que anteceden al "estado de cosas satisfactorio". Este último término al poco tiempo se convertiría en el concepto que hoy conocemos como "refuerzos" y que en estas notas llamaremos también SBI. En lo que resta del capítulo, revisaremos la historia y la evidencia acerca del papel de la contigüidad en la asignación de crédito para una respuesta.

En 1947, Skinner publicó los resultados de un pequeño experimento para demostrar la suficiencia de la contigüidad para el aprendizaje de respuestas. A las palomas hambrientas se les presentó comida cada 15 seg., independientemente de su comportamiento. Se observó que a pesar de esto, muchas palomas desarrollaron comportamientos estereotipados, como girar en círculos o picotear ciertas áreas. Los comportamientos difieren de paloma a paloma. (Ver figura). Skinner explicó estos resultados, señalando que para cada ave, una respuesta ocurría de forma accidental inmediatamente antes del refuerzo y esa contigüidad era responsable del fortalecimiento de dicha respuesta. A partir de este tipo de observaciones, Skinner concluyó: "Decir que un reforzador es contingente sobre una respuesta no significa otra cosa que decir que "se presenta después de la respuesta". Para Skinner, presumiblemente, el condicionamiento ocurre únicamente debido a la relación temporal, expresada en términos de la proximidad entre la respuesta y el reforzador.

La sencilla historia anterior fue rápidamente cuestionada en una réplica del es-

tudio de Skinner, publicada por Staddon y Simmelhag en 1971. Al igual que Skinner, a un grupo de palomas se le dió acceso a comida cada 15 segundos, independientemente de su comportamiento. A diferencia de Skinner, estos investigadores observaron cuidadosamente el comportamiento de las palomas a lo largo de los 15 segundos. Los resultados se muestran en la figura x. No encontraron evidencia de que se aprendiera la respuesta individual que accidentalmente antecede el acceso a la comida. No obstante, observaron que para todas las palomas, el comportamiento desplegado se podía agrupar en dos clases: una de respuestas que ocurrían al final del intervalo, a las que llamaron "respuestas terminales", y las cuales incluían, entre otras, el orientarse hacia la pared del comedero; y una segunda clase de respuestas que agrupaba comportamientos que ocurrían a la mitad del intervalo, a las que llamaron "respuestas interinas", entre las cuales se observó la conducta de picar el piso. El estudio se replicó con ratas y en ese caso también se observó la agrupación de respuestas en dos clases.

De los resultados del experimento de Staddon y Simmelhag pueden extraerse dos conclusiones, una negativa y otra positiva. Primero, en relación al tema de este capítulo, podemos concluir que en los experimentos en los que no existe una relación netamente causal entre respuesta y SBI, el estímulo no selecciona a la respuesta que accidentalmente le antecede: contradiciendo a la idea que la contigüidad es una condición suficiente para el aprendizaje de respuestas. La conclusión positiva es que la mera presentación de un SBI induce un conjunto tipificado de respuestas, y que la periodicidad de la presentación del SBI organiza el comportamiento de los organismos alrededor del tiempo. En otro capítulo revisaremos en detalle otros resultados relacionados y su papel en una teoría general del comportamiento.

6.0.1 ¿Es la contigüidad entre una respuesta y un refuerzo una condición necesaria para la adquisición de la respuesta?

En el experimento de superstición de Skinner no había una relación de dependencia causal entre respuesta y refuerzo y Skinner buscaba demostrar que la mera contigüidad era suficiente para el aprendizaje de respuestas. Pero esta demostración dependía de la manera en la que se especifica el concepto de contigüidad en términos concretos (¿cuándo podemos considerar que un suceso es realmente contiguo? Si el SBI ocurre un segundo antes del EC o dos segundos antes, ¿sigue siendo contiguo?¿cuándo comienza y cuándo termina la contigüidad?). Para poder estudiar sistemáticamente el papel de la contigüidad, es necesario especificar la ventana temporal que define a dos eventos como contiguos. La estrategia teórica-experimental inicial en esta área fue considerar como contigüidad una ventana de cero segundos y considerar el impacto de ventanas mayores como distintas instancias de efectos de la demora en el refuerzo.

Para poder analizar el efecto de diferentes demoras, se requiere poder controlar

esa relación. Con esa finalidad, es necesario estudiar protocolos en los que exista una relación de dependencia entre la respuesta y el SBI, en particular, protocolos en los cuales se varíe el tiempo entre las respuestas y la presentación de los SBI que son generados por estas respuestas. A continuación revisaremos el efecto de variar el valor temporal de los intervalos entre respuestas y los SBI producidos por ellas.

En un primer experimento, Dickinson y sus colaboradores evaluaron el impacto de diferentes demoras entre la respuesta de apretar una palanca y la presentación del SBI. El experimento se condujo con ratas sin ninguna experiencia previa con el procedimiento. El propósito del experimento fue evaluar si las ratas aprenderían a apretar la palanca. Cada respuesta producía un SBI con una demora fija. Durante el periodo de demora, las ratas podían volver a responder y producir otro SBI con una demora igual. Noten que con este procedimiento podía darse el caso de que accidentalmente una de las respuestas de las ratas durante uno de los periodos de demora ocurriera justo antes del SBI. Para descartar el aprendizaje de respuestas por la mera contigüidad accidental con el SBI (esto es, en ausencia de un efecto causal sobre este) había que evaluar si el aprendizaje de la respuesta en las ratas se genera independientemente de la existencia de una dependencia causal entre respuesta y refuerzo. Para descartar esta posibilidad, se estableció un grupo diferente de ratas a las cuales se les entregaba el SBI al mismo tiempo que lo recibía el grupo dependiente: para este segundo grupo, sus respuestas no tenían ningún efecto sobre el momento de aparición de los reforzadores. Si ambos grupos mostraban los mismos patrones de respuesta, eso significaba que la contigüidad temporal era el factor determinante del aprendizaje de las respuestas y que las ratas del primer grupo no aprendían en función del poder causal que identificaban en sus respuestas. Esto también implicaría que las ratas del primer grupo, el grupo dependiente, estarían aprendiendo a asociar la presentación de los SBI con sus respuestas que se generaban accidental y tardíamente dentro del intervalo de demora. La figura x muestra los resultados del experimento. Se probaron tres valores de demora adicional a la condición de contigüidad estricta. La medida empleada para determinar si se había aprendido la respuesta fue el número de respuestas de apretar la palanca por minuto. En el panel izquierdo puede verse que la tasa de respuestas va decreciendo conforme incrementa la demora hasta alcanzar un valor de 32 segundos entre respuesta y reforzador, y cero aprendizaje con demora de 64 segundos. En el panel de la derecha, se muestra el efecto de la dependencia respuesta-SBI sobre el aprendizaje de las respuestas de las ratas a lo largo de 20 sesiones. Se compara la tasa de respuesta para el grupo con dependencia respuesta-SBI con la del grupo que recibía el SBI independientemente de las respuestas de las ratas. Los resultados indican que para el segundo grupo de ratas, la tasa de respuesta fue casi cero para todas las demoras evaluadas, es decir no hubo aprendizaje, mientras que para el primer grupo de ratas se observaron los patrones de aprendizaje con demoras descritos previamente.

Como ya se mencionó, un problema a resolver para el experimento anterior era controlar la posibilidad de que los organismos aprendieran respuestas que

aparecieran accidentalmente contiguas al SBI. Lattal y Gleeson realizaron un experimento con una ingeniosa estrategia alterna para controlar esta posibilidad. En su experimento, la respuesta de las palomas de picar una tecla detonaba una demora de 10 seg. después de la cual obtenían un refuerzo de comida; sin embargo, esto último sólo ocurría si las palomas no daban ninguna respuesta durante el periodo de demora. De esta forma, el diseño experimental garantizaba una demora real de 10 segundos. En la Figura x, puede verse que aún eliminando la posibilidad de una contigüidad accidental, las palomas aprenden la respuesta de picar la tecla.

6.0.2 Percepción de la relación de causalidad respuesta - refuerzo

Los resultados que hemos reportado en este capítulo, nos llevan a considerar la siguiente pregunta: ¿Pueden los organismos discriminar entre refuerzos producidos por su comportamiento de aquellos que son independientes de él?

Con un ingenioso experimento, Killeen pretendió explorar esta pregunta de manera directa. En el experimento, la tarea para las palomas era discriminar si un cambio en la iluminación de una tecla era el resultado o no de su comportamiento. La respuesta de picar una tecla central iluminada tenía como consecuencia el que, de manera aleatoria, cinco de cada 100 respuestas (probabilidad de 0.05) causaran que se apagara la tecla central y se encendieran dos teclas laterales. A la tasa que la paloma pica la tecla iluminada, la computadora generaba pseudo picotazos que tenían también una probabilidad de 0.05 de apagar la luz de la tecla central y encender las luces de dos teclas laterales. La tarea para las palomas era discriminar si el apagado de la tecla central era consecuencia de uno de sus picotazos o de los producidos por la computadora, es decir los apagados independientes de su respuesta. Si el apagado dependía de la respuesta de la paloma, entonces la respuesta de picar la tecla derecha permitía acceso a la comida; por otro lado, si el apagado era independiente de su respuesta, entonces el picar la tecla izquierda era la respuesta que producía acceso a la comida. Los errores tenían como consecuencia el apagado de todas las luces por un breve periodo de tiempo. La forma en la cual la paloma informaba sobre su juicio era por sus respuestas a las teclas laterales.

Como veremos en la práctica sobre "teoría de detección de señales", cuando el organismo identifica correctamente al cambio en la tecla que es dependiente de su respuesta le llamamos un "hit"; por el contrario, cuando el organismo identifica al cambio en la tecla como dependiente de su respuesta cuando en realidad era el resultado de una pseudo respuesta, le llamamos una "falsa alarma". En la práctica sobre teoría de la detección de señales (poner link), veremos que las respuestas de los animales no dependen exclusivamente de su capacidad para detectar causalidad. En este contexto es que emerge la siguiente pregunta: ¿Cuando los organismos muestran alguna respuesta "supersticiosa", esta se debe a una falla del mecanismo de discriminación o a las ganancias o costos ligados

a emitir una respuesta causalmente errónea? Ponderen que harían ustedes si un hit produjera $10, y una falsa alarma les restara un peso. Ahora consideren que harían si los hits les otorgaran $1,000 y las falsas alarmas siguieran teniendo un costo de un peso. Incrementen ahora la ganancia para los hits a $10,000. Seguramente, conforme la ganancia para los hits fuese incrementando, su estrategia se iría acercando a responder con mayor frecuencia que ustedes produjeron el cambio, aunque la probabilidad real de que esta relación causal sea verdadera se mantiene inalterada. En el experimento, Killeen varió la cantidad de comida que las palomas recibían después de un hit y encontró que las palomas se comportan justo como lo haríamos nosotros.

¿Qué papel juega el tiempo transcurrido entre una respuesta y el encuentro con un refuerzo independiente? ¿Si ha transcurrido un tiempo largo entre una respuesta del organismo y un evento accidental, este todavía le asignará poder causal a su respuesta para dar cuenta de la ocurrencia del evento? En el mismo experimento, Killeen se preguntó acerca del efecto del tiempo transcurrido entre una respuesta y el apagado de la luz generado por la computadora sobre la probabilidad de una falsa alarma. Encontró que las falsas alarmas se reducen conforme ese tiempo se mueve de 0.20 seg. a 1.0 seg. Ver fig. Es decir, a mayor distancia entre la respuesta y el suceso accidental, existe una menor probabilidad de que el organismo le asigne poder causal a su respuesta para explicar la ocurrencia del evento accidental.

6.0.2.1 Conclusiones

De los experimentos presentados podemos alcanzar las siguientes conclusiones:

1. La evolución ha seleccionado mecanismos de aprendizaje que consisten en buscar los mejores predictores de los sucesos biológicamente importantes (SBI).
2. Existen sesgos para reducir el tamaño del espacio de estímulos/respuestas candidatos predictores de SBI.
3. Existen sesgos como la relevancia biológica, la novedad y la contigüidad.
4. La contigüidad es un factor que influye la selección de estímulos/respuestas candidato, pero no es una condición ni necesaria ni suficiente para el aprendizaje.
5. En los casos en los que se considera a compuestos de estímulos como conjuntos de elementos, parece haber competencia entre estos por la asignación de crédito para la predicción del SBI.
6. El crédito asignado previamente a un elemento del compuesto le resta (bloquea) la posibilidad de asignar crédito a otro elemento. Lo anterior significa que no todos los elementos contiguos al SBI necesariamente son considerados como predictores del mismo.

7. La contigüidad estricta entre respuesta y SBI tampoco es necesaria para el aprendizaje de respuestas. Las palomas y las ratas pueden aprender una respuesta aún con demoras de 32 segundos.

8. La mera presentación del refuerzo, independiente de la respuesta, no es suficiente para generar el aprendizaje de una respuesta.

9. La contigüidad estricta no es necesaria para el aprendizaje de respuestas: pero mientras más cercano esté el refuerzo de una respuesta, más fácil es su adquisición.

10. Aún en protocolos en los cuales no es posible la contigüidad accidental entre respuesta y SBI, los animales adquieren la respuesta que genera reforzadores demorados en el tiempo.

11. Las palomas pueden discriminar entre consecuencias dependientes e independientes de su respuesta.

12. El que un organismo juzgue a una consecuencia como dependiente de su respuesta varía en función de la ganancia y el costo asociado a esos juicios.

Chapter 7

Correlación, Tiempo y Contingencia

En mi juventud, los periódicos más amarillistas de la Ciudad de México tenían titulares de ocho columnas, del tipo "mariguano ataca con un cuchillo a su vecino". Titulares de este tipo eran utilizados por los noticieros para justificar la prohibición del consumo de la mariguana. Independientemente del debate acerca de su legalización, esos titulares tienen un importante problema que dificulta atribuir al consumo de la droga el ataque perpetuado. El error radica en asignar el crédito del ataque a la mariguana, observando únicamente lo que sucede después del consumo de la droga. Los periódicos y los noticieros debían preguntarse adicionalmente cuántos ataques a vecinos ocurren cuando el atacante **No** está bajo la influencia de la droga. Si el número de ataques a vecinos fuese similar cuando el atacante consumió la droga y cuando este no lo hizo, consideraríamos que el consumo de esta sustancia no se encuentra correlacionado con los ataques. De la misma manera, nos planteamos las dos preguntas sutilmente diferentes pero relacionadas de: ¿cuántos vecinos **No** fueron atacados cuando se encontraron con una persona que **Sí** había consumido la droga? y ¿cuántos vecinos **No** fueron atacados cuando se encontraron con una persona que **No** la había consumido? Si el número de ataques y de no ataques a vecinos en encuentros con otras personas es similar, independientemente de si las personas han consumido la droga o no, consideraríamos que su consumo no está correlacionado con los ataques.

En este capítulo, nos preguntaremos si las palomas y las ratas son sensibles únicamente a lo que ocurre después de un estímulo o de una respuesta o si para la asignación de crédito, estos organismos también contemplan lo que ocurre cuando el estímulo o la respuesta en cuestión no se encuentran presentes. En otras palabras, ahondaremos con mayor profundidad sobre la pregunta de si la contigüidad entre un estímulo (o una respuesta) y un reforzador es un elemento

suficiente y necesario para la asignación de crédito.

Los protocolos experimentales presentados en los dos capítulos anteriores tienen en común el variar lo que ocurre después de una estímulo o de una respuesta, ya sea con un refuerzo inmediato o demorado. En 1968, Rescorla introdujo un protocolo experimental que permite manipular lo que ocurre en la presencia y en la ausencia de un estímulo o una respuesta. Al protocolo le llamó *verdaderamente aleatorio*. En lugar de presentar un refuerzo al final del estímulo condicionado (EC), Rescorla varió la probabilidad de un refuerzo durante la presencia y durante la ausencia del estímulo condicionado (ver figura). En su experimento, Rescorla presentó un estímulo condicionado de una duración de 20 seg. con un intervalo de 2 min. entre cada presentación del EC: a este último periodo se le conoce como el intervalo entre ensayos. Durante ambos periodos de tiempo, la presentación del refuerzo se determina con cierta probabilidad para cada segundo. Las diferencias entre las probabilidades en los dos periodos de tiempo pueden manipularse, de tal forma que la probabilidad de un refuerzo durante la presencia del EC sea mayor o menor que la probabilidad durante el intervalo entre ensayos. En el primer caso, diremos que el EC y el refuerzo están correlacionados positivamente, en el segundo caso diremos que están correlacionados negativamente. Si las probabilidades en ambas duraciones son iguales diremos que ambos elementos no se encuentran correlacionados. La figura x ilustra el espacio de posibles correlaciones. En el eje de las X se presenta la probabilidad de refuerzo durante el intervalo entre ensayos, mientras en el eje de las Y se presenta la probabilidad de refuerzo durante el estímulo condicionado. La línea diagonal representa la falta de correlación entre el EC y el refuerzo. El espacio arriba de la diagonal representa correlaciones positivas y el espacio abajo de la diagonal representa correlaciones negativas. Los puntos cercanos al 1 y al cero, representan las correlaciones más fuertes: en un caso, los refuerzos ocurren exclusivamente durante el EC, en el otro, ocurren casi exclusivamente durante el intervalo entre ensayos.

Usando diferentes medidas de aprendizaje, múltiples experimentos han encontrado que en la condición *no correlacionada* los animales no le asignan crédito al EC. En cambio, cuando la correlación es positiva, los animales aprenden que el EC predice al refuerzo, y por el contrario, cuando la correlación es negativa, los animales aprenden que el EC predice la ausencia del refuerzo. Esos resultados añaden evidencia a la afirmación de que la contigüidad entre el EC y el refuerzo no es una condición suficiente para la asignación de crédito. Cuando es igualmente probable que el refuerzo aparezca en la presencia del EC como en su ausencia, en el entorno del organismo no hay una relación causal entre el EC y el refuerzo. Intuitivamente, en la condición no correlacionada, el EC no proporciona información alguna acerca de la ocurrencia de los refuerzos. Podríamos eliminar del protocolo experimental el EC y no se alteraría la expectativa del organismo acerca de la ocurrencia de los refuerzos. En otro capítulo, veremos que formalizar la noción intuitiva de información ilumina muchos de los hallazgos acerca de la asignación de crédito.

Los organismos no tan solo son sensibles a la correlación entre estímulos y refuerzos, a continuación veremos que también son sensibles a la correlación entre respuestas y refuerzos. Para evaluar si la contigüidad entre la respuesta y el refuerzo es suficiente para asignar crédito o si es necesario que exista una correlación entre estos elementos, Hammond corrió un experimento con una estructura similar al de Rescorla, manipulando la probabilidad de un refuerzo dada la ocurrencia y ausencia de una respuesta. Para lograr igualar la oportunidad de una respuesta y de una no respuesta, Hammond partió el tiempo de la sesión experimental en segmentos de un segundo de duración. En cada uno de esos segundos el animal puede o no emitir una respuesta; asimismo, este puede obtener o no, con cierta probabilidad, un refuerzo. La probabilidad de obtener un refuerzo en cada segundo dependía de la presencia o ausencia de una respuesta por parte del organismo. En todas las condiciones del experimento, Hammond mantuvo constante la probabilidad de un refuerzo dada la ocurrencia de una respuesta en el intervalo de un segundo, y varió su probabilidad en la ausencia de la respuesta. La probabilidad del refuerzo dada la respuesta fue de 0.05, mientras que la probabilidad de un refuerzo dada la no respuesta fue de cero o de 0.05. La figura x muestra que las ratas no responden en las condiciones experimentales en las cuales las dos probabilidades de refuerzo eran iguales.

7.1 Tiempo

Un problema de dirigir nuestra atención exclusivamente al momento en el que ocurre el estímulo condicionado (EC) o la respuesta, es que solo atendemos a los refuerzos contiguos al EC; en ese sentido, también emerge el problema de ignorar el tiempo entre presentaciones del EC, lo que llamamos el intervalo entre ensayos (TEE). Si la contigüidad estímulo-respuesta fuera la única variable importante para el aprendizaje de comportamientos, entonces manipular la duración del EC relativa a la duración del intervalo entre ensayos no tendría ningún efecto sobre el aprendizaje.

Un escenario hipotético nos hace dudar de la conclusión anterior. Comparemos dos fábricas, en ambas, cada cuatro horas hay 10 minutos de descanso. Pero en una de las fábricas, el período de descanso es precedido por una señal que dura 3 horas y 45 minutos, mientras que en la otra, el descanso es precedido por una señal que dura 10 minutos. Pregúntense si ambas señales les serían igualmente informativas, si les prestarían igual atención a ambas o si una de ellas les permitiría anticipar adaptativamente la ocurrencia del descanso. Intuitivamente, este caso hipotético nos hace pensar que los escenarios con EC con una duración muy larga respecto a la duración de los intervalos entre ensayos (TEE) inducen un menor aprendizaje que los escenarios con EC que tienen una duración más corta respecto a la duración de los TEE.

Gibbon et al. llevaron a cabo justo ese protocolo con ratas para evaluar la importancia de las duraciones del intervalo entre ensayos y del estímulo condicionado.

En el experimento, se manipularon dos condiciones: en la primera, se incrementó la duración del estímulo condicionado, manteniendo constante el tiempo del intervalo entre ensayos; y en la segunda condición, se incrementaron ambas duraciones proporcionalmente (por ejemplo, de 4 a 8 y de 48 a 96 segundos, respectivamente). La medida del aprendizaje fue el número de refuerzos necesarios para mostrar aprendizaje. La siguiente figura muestra los resultados. La línea roja muestra que el número de refuerzos requeridos para el aprendizaje incrementó como función del aumento en la duración del EC, manteniendo constante la duración del intervalo entre ensayos: bajo esta condición, las señales fueron "menos informativas" para los organismos, por lo que el aprendizaje resultó "más difícil". Por el contrario, incrementar las dos duraciones, manteniendo constante su razón, no tuvo un efecto sobre el número de refuerzos necesarios para el aprendizaje.

Lo anterior indica que los organismos son sensibles a la razón TEE/TEC: si esta razón rebasa un valor, el animal aprende acerca de la importancia del EC. Una forma de entenderlo es suponer que el EC reduce la incertidumbre acerca del momento de ocurrencia de un refuerzo: mientras la razón TEE/TEC sea más grande, mayor es la reducción en la incertidumbre de la entrega. En otras palabras, la asignación de crédito a un estímulo o respuesta depende de que el estímulo condicionado sea breve relativo al intervalo entre las presentaciones del refuerzo. En nuestro ejemplo hipotético, el obrero de la fábrica aprenderá acerca de la señal de 10 minutos que ocurre cada 4 horas, y no de la señal de 3 horas y media de duración.

Conclusiones

Adicionalmente a las limitaciones ya señaladas sobre el papel de la contigüidad en la selección de estímulos/respuestas candidato, cabe agregar las siguientes observaciones:

1. La correlación en el tiempo entre EC y EI es un factor importante en el aprendizaje.
2. Uno de los criterios para identificar un buen predictor de SBI es su duración relativa a la duración del intervalo entre ensayos: en particular, la razón TEE/TEC.

Chapter 8

Modelo de Aprendizaje por Refuerzo

Como un resultado de la selección natural, los agentes biológicos son sistemas con mecanismos que les permiten detectar, predecir y controlar los sucesos biológicamente importantes (SBI). Recordemos que los SBI son sucesos con valor hedónico, también conocidos como refuerzos, los cuales incrementan el éxito reproductivo diferencial de los organismos. Para ejecutar estas funciones de detección, predicción y control de SBI los organismos necesitan hacer contacto con la estructura causal del entorno: estructura a la cual designamos bajo el término de *propiedades estadísticas del entorno*.

Una parte de la estructura estadística de los SBI consiste de estímulos que se despliegan en el tiempo y en el espacio, los cuales en algunas ocasiones aparecen solos, mientras que en otras instancias aparecen seguidos (contiguos) de un refuerzo. Hay tiempos sin nubes y otros con nubes, estos últimos pueden ser seguidos o no de lluvia. La tarea para el organismo es determinar si existe una relación causal entre las nubes y la lluvia. Una segunda parte de la estructura estadística del entorno consiste del hecho de que, con relación al despliegue de respuestas de los organismos en el tiempo y en el espacio, algunas ocasiones estas respuestas van seguidas de un refuerzo y en otras ocasiones no. Hay veces en las que ustedes le dicen "hola" a una persona y otras en las que no. Después del "hola" ustedes pueden recibir, o no, otro "hola" de regreso. Los protocolos de condicionamiento clásico e instrumental definen, como una primera aproximación, las estructuras causales más simples que se estudiaron en buena parte del siglo XX. En los dos capítulos anteriores, resumimos la literatura empírica acerca del papel de la contigüidad en la asignación de crédito a estímulos y respuestas que permiten *predecir y controlar* los refuerzos. Presentamos evidencia que muestra que aunque la contigüidad es importante, esta no es ni necesaria, ni suficiente para la asignación de crédito.

Para dar cuenta de los resultados empíricos presentados en los capítulos anteriores, en la segunda mitad del siglo XX se consolidó un modelo matemático, conocido como *Aprendizaje por Refuerzo*, junto con varias modificaciones del mismo. El propósito de este capítulo es presentar estos modelos y su desarrollo como respuesta a la evidencia sobre el papel de la contigüidad.

.

8.0.1 Curvas de Aprendizaje

El aprendizaje es un proceso dinámico, que describe los cambios en el comportamiento como una función de la experiencia de los organismos. Desde finales del siglo XIX se han obtenido "curvas de Aprendizaje" que describen cambios en alguna medida de ejecución de los organismos como una función de las ocasiones en las que los estímulos que encaran o las respuestas que despliegan van seguidos de un refuerzo.

Como ilustración, presentamos dos ejemplos, el primero es la curva de adquisición y extinción de la frecuencia del reflejo de parpadeo, en un protocolo de condicionamiento clásico con humanos. Las curvas representan diferentes intensidades del soplo al ojo. (Parssey, 1948)

IMAGEN

Un segundo ejemplo es la curva de adquisición de la velocidad de tocar una palanca en un protocolo de condicionamiento instrumental con ratas (Ramond, 1954). Las dos curvas representan los datos obtenidos con diferentes niveles de privación.

IMAGEN

La siguiente curva de adquisición idealizada, captura los datos de adquisición presentados en las dos figuras anteriores. Podemos ver que es una curva negativamente acelerada de ganancias decrecientes, esto es, el impacto de que un refuerzo siga a una respuesta se va reduciendo conforme se acumulan las ocasiones en las que una respuesta va seguida de un refuerzo.

IMAGEN

Nota. Una alternativa teórica supone que el aprendizaje no es un proceso gradual, sino uno de cambio abrupto. Si este fuese el caso, las curvas de adquisición de crecimiento gradual decreciente podrían ser un artefacto de promediar la ejecución de animales con cambios *no* continuos en el aprendizaje. Considere el caso de un grupo de animales cuya ejecución es promediada. Uno de los animales empieza a responder al ensayo 10, otro al 20, otro al 30, otro al 40 y así hasta un animal que responde al ensayo 80: al promediar estos datos, la curva de aprendizaje aparecerá como una curva continua. Por esta razón, Skinner, Estes, Spence y más recientemente Gallistel argumentan en favor de analizar los datos de sujetos individuales y, preferentemente, registros acumulativos en lugar de

los datos promediados. En ese mismo sentido, los modelos que presentaremos a continuación resultan válidos cuando los datos analizados provienen de sujetos individuales.

Los modelos de refuerzo modelan la forma de las curvas de adquisición obtenidas empíricamente. El modelo de refuerzo más general es un sistema dinámico que describe los cambios en el valor de un estímulo y/o respuesta a lo largo del tiempo, como una función del número de ocasiones en las que un estímulo o una respuesta van seguidas de un refuerzo.

Pasen al simulador de ecuaciones en diferencia para entender los modelos dinámicos discretos.

8.1 Modelo de Refuerzo

Los modelos de refuerzo le asignan un número a cada estímulo y/o respuesta: esta magnitud representa la calidad del estímulo/respuesta como predictor de un refuerzo. A lo largo del siglo XX, a esta magnitud se le conoció como fuerza del reflejo, fuerza del hábito, fuerza asociativa del estímulo y fuerza de la respuesta. Para nuestros propósitos el número refleja el *valor* predictivo de un estímulo o de una respuesta y simplemente le llamaremos el valor V del estímulo i, o el valor Q de la respuesta i.

El modelo de refuerzo propone que después de cada presentación de un estímulo o la emisión de una respuesta, su valor se actualiza como una función de si este va seguido o no de un refuerzo. Es importante señalar que el modelo asume que la actualización del valor de un estímulo o una respuesta solo ocurre en las ocasiones en las que estos se presentan. El valor predictivo de un tono solo se actualiza cuando el tono se presenta y no en su ausencia. Esto implica que el mero paso del tiempo sin el tono o la presentación de otros estímulos no alteran el valor del tono. Por esta razón, a estos modelos se les llama también modelos basados en ensayos. Esto es, la variable importante que determina el valor predictivo (la asignación de crédito) es el número de ocasiones en las que un estímulo o una respuesta son seguidos de un reforzador. Mientras mayor es este número, mayor será el valor adquirido por el estímulo o la respuesta. De forma complementaria, el valor de estos estímulos o respuestas decrementa cuando estos se presentan sin ser seguidos por el reforzador.

En 1950, Bush y Mosteller propusieron una versión matemática del modelo de aprendizaje por refuerzo esbozado en el párrafo anterior. Esta versión sigue siendo la base que sustenta varios de los modelos más recientes, tanto en la psicología como en el aprendizaje de máquinas.

La estructura matemática de los modelos de refuerzo tiene diferentes interpretaciones teóricas. Nosotros consideraremos dos:

1. Un proceso de carga - descarga y

2. Un proceso de reducción del error

8.1.1 Proceso de carga - descarga

Los modelos de refuerzo son una propuesta de solución computacional a la asignación de crédito. La solución incluye dos pasos: el primero es la reducción del tamaño inicial del espacio de candidatos para incluir únicamente sucesos que son contiguos, similares, novedosos y evolutivamente relevantes con relación a los SBI. El segundo es un mecanismo que permita ir reduciendo, a través de la experiencia, el espacio de candidatos hasta terminar con uno solo. El modelo de refuerzo canónico combina un algoritmo de ascenso de colina con el sesgo de contigüidad. Bush y Mosteller (1951) formalizaron esta clase de modelos que, en diferentes variantes, han dominado la literatura teórica y experimental en el estudio del aprendizaje a partir de la década de los 70s del siglo pasado.

Como ya se dijo antes, la forma más literal de interpretar el modelo de refuerzo es como un proceso en el cual: cada refuerzo incrementa (carga, fortalece) el valor del estímulo / respuesta que le antecede y cada ocurrencia del estímulo / respuesta sin ser acompañado de un refuerzo decrementa (descarga) su valor. Este es un proceso en el que la variable V se actualiza con cada ocurrencia del estímulo / respuesta como una función que varía dependiendo de si el estímulo/respuesta va seguida o no de un refuerzo. La carga de la batería de su celular es un ejemplo muy cercano a su vida cotidiana que ejemplifica el proceso descrito. Para que su celular funcione, ustedes tienen que conectarlo a una toma de corriente. Mientras está conectado, la carga de la batería se va actualizando hasta llegar a un punto máximo. Al usarlo sin tenerlo conectado, la batería se va descargando como una función del uso del celular sin una carga adicional.

Veamos cómo se aplica este modelo en el caso de un protocolo estándar de condicionamiento clásico: en este contexto, se observa un estímulo/respuesta candidato a la asignación de crédito y a cada presentación de él se le conoce como un ensayo. Cada ocurrencia de un ensayo puede estar acompañado o no de un SBI. Consideremos que Vx represente el valor predictivo de un estímulo x . Nos interesa la dinámica del cambio en Vx conforme un organismo experimenta ensayos en los que un estímulo condicionado x va seguido de un refuerzo. Para facilitar la aplicación del modelo, supondremos que el tiempo es discreto y el subíndice t representa el momento en que se presenta el estímulo x. La variable $Vx_t + 1$ es el valor actualizado del estímulo x en el siguiente ensayo $t + 1$. La variable R representa si se presentó o no un refuerzo después del estímulo x . R puede tener solo dos valores, uno si el refuerzo se presenta después del estímulo x o cero si el estímulo x se presenta sin ser seguido por el refuerzo.

Vamos a asumir que Vx_t+1 depende sólo de dos factores: 1. su valor acumulado hasta el ensayo inmediatamente anterior Vx_t 2. el valor de R_t en el ensayo actual

$$Vx_{t+1} = Vx_t + R$$

3. La expresión anterior supone que el efecto del valor de V en el ensayo anterior tiene el mismo peso que el reforzador presentado o no en el momento actual. Nosotros deseamos que la ecuación capture la importancia relativa de las dos variables. Para ello, supondremos que el impacto de esas dos variables es una suma ponderada, donde el parámetro a representa el peso de ponderación asignado a cada uno de los dos factores. Si el valor de a esta entre cero y uno, los parámetros asociados con cada factor serán a y $(1-a)$.

$$Vx_{t+1} = (1-a)Vx_t + aR_t$$

donde: $0 < a < 1$.

La ecuación anterior es una ecuación recurrente, en la que en cada iteración (presentación del estímulo x) hay siempre un Vx viejo y un Vx nuevo. La ecuación describe la actualización de Vx de ensayo a ensayo, como una función del valor de Vx acumulado hasta el ensayo anterior y la presentación o ausencia del SBI. Vx_t es la integración, el acumulado, de todas las experiencias previas con el refuerzo. En cada ensayo, la Vx nueva del ensayo anterior se convierte en la Vx vieja del presente ensayo, contribuyendo a generar una nueva Vx. Pase al simulador x para revisar ecuaciones recurrentes.

El parámetro a, que multiplica al refuerzo, determina la velocidad del aprendizaje: mientras mayor sea su valor, más rápido será el aprendizaje. Una forma de entender el papel de a, es considerarlo como el parámetro que especifica la importancia de la experiencia acumulada hasta el momento (el valor del pasado), relativa a la ocurrencia o no de un refuerzo (el valor del presente). Valores cercanos a cero sugieren que la experiencia acumulada es más importante que una nueva experiencia con un refuerzo, resultando en poco aprendizaje, mientras que valores cercanos a uno sugieren que la presentación del refuerzo es más importante que la experiencia acumulada hasta ese momento, resultando en un rápido aprendizaje. Consideren una interacción de larga duración con una amistad que ha resultado en un valor alto para ustedes asociado con esa relación. Nos podemos preguntar cuál es el impacto de que una mañana esa amistad no los salude. Si el parámetro a fuese cercano a cero, el no saludo no modificaría sustancialmente el valor V de la amistad, mientras que si el valor de a fuese cercano a uno, a pesar de los años de experiencias positivas con esa amistad, su impacto en el valor V sería más significativo.

Las siguientes figuras muestran los resultados, ensayo a ensayo, de una simulación con diferentes valores del parámetro a.

IMAGENES

8.1.2 Reducción del error de predicción como motor del aprendizaje

El reacomodo de los términos de la ecuación de carga - descarga permite una interpretación alternativa del modelo de aprendizaje por refuerzo: esta vez en términos de un mecanismo de reducción en el error de predicción. Estos modelos representan hoy la versión dominante en la psicología del aprendizaje y las neurociencias.

Arreglando los términos de la ecuación del integrador con fuga:

$$V_{t+1} = (1 - a)V_t + aR_t$$

$$V_{t+1} = V_t - aV_t + aR_t$$

$$V_{t+1} = V_t + a(R_t - V_t)$$

Restándole V_t a ambos lados de la ecuación anterior, dejando que $\Delta Vx = Vx_t +$ 1–Vx_t entonces el cambio de ensayo a ensayo es descrito, agrupando términos:

$$\Delta Vx = V_t + a(R_t - V_t) - V_t$$

$$\Delta Vx = a(R_t – Vx_t)$$

Esta segunda forma de la ecuación, enfatiza la magnitud del cambio momento a momento, en lugar del valor del estímulo momento a momento.

En cualquiera de las dos formas de presentar la ecuación del integrador, el valor V es una función de la diferencia entre el refuerzo obtenido y el valor del estímulo en el tiempo t. A esta diferencia dentro del paréntesis se le conoce como el error de predicción: la diferencia entre la R que se obtiene y lo que se esperaba obtener, V. Cuando esta diferencia es igual a cero, no habrá cambios en el valor de V. Es por esta forma de la ecuación que estamos llamando a V_t el valor predictivo de la respuesta. El parámetro a, entre 0 y 1, sigue representando la velocidad del aprendizaje, en este caso, la importancia del error de predicción.

En la literatura contemporánea, a la ecuación anterior se le conoce como regla delta.

$$V_{t+1} = V_t + a\Delta Vx$$

donde delta = (lambda_t - Vx_t) es el error de predicción, es decir, el motor del aprendizaje. Bajo el formato que enfatiza la magnitud del cambio, delta se incorpora a la ecuación de la siguiente forma:

Vx = a(delta)

Chapter 9

El Modelo de Rescorla y Wagner

En el capítulo anterior, vimos que el modelo de aprendizaje por refuerzo, conocido también como modelo de *aprendizaje de error de predicción* captura razonablemente bien la adquisición de valor predictivo cuando un estímulo o respuesta van seguidos de un refuerzo. En este capítulo veremos que este modelo *no* puede dar cuenta de los resultados presentados en el capítulo sobre asignación de crédito, los cuales ilustran la importancia de una serie de correlaciones: aquellas que sostiene el EC o la respuesta con la aparición del refuerzo; aquellas que sostiene el EC con otros estímulos presentes (elementos del contexto); y aquellas que existen previamente entre otros estímulos distintos y el refuerzo actual. Originalmente, estos últimos resultados indujeron interpretaciones que enfatizaban que la asignación de crédito a un estímulo o respuesta dependía de que estos fueran seguidos de un refuerzo que era *sorpresivo, inesperado, informativo, o que atraía la atención.* En 1972, Rescorla y Wagner presentaron un modelo que daba cuenta de los resultados que muestran que la mera contigüidad no es un factor necesario ni suficiente para la asignación de crédito. El modelo es una extensión del principio de la reducción de error, que captura la intuición acerca del papel de la sorpresa como un modelo matemático: todo ello sin hacer referencia a procesos atencionales que se suponían difíciles de evaluar con sujetos no humanos. Este modelo sigue siendo hasta la fecha el motor de la investigación en aprendizaje.

9.0.1 Modelo de Rescorla y Wagner

El modelo de Rescorla y Wagner incluye dos grandes componentes. El primer componente es el *Modelo de Refuerzo* de Bush y Mosteller, el cual hemos visto que establece la reducción en el error de predicción como el motor del apren-

dizaje. El segundo componente es un modelo de la forma en la que un organismo percibe estímulos compuestos. En particular, este modelo supone que los organismos perciben a los estímulos, por ejemplo un rostro, como un conjunto de elementos separables: en este caso, un rostro se percibe en términos de ojos, nariz, labios, entre otros. El modelo asume que todos estos elementos compiten entre ellos por la asignación de crédito.

El modelo de aprendizaje utilizado por Rescorla y Wagner es una variante del modelo de la reducción en el error de la predicción (también conocido como la regla delta).

$$Vx_{t+1} = Vx_t + a(R_t - Vx_t)$$

Donde 0 < a <1

Para entender el segundo componente del modelo de Rescorla y Wagner, consideremos por un momento las características del entorno modelado. Hasta antes de los años 60s del siglo pasado, los investigadores limitaban sus experimentos a protocolos en los que se presentaba un solo estímulo condicionado. Sin embargo, los entornos reales no consisten de elementos que aparecen aisladamente y cuyo único aspecto complejo es la variabilidad en su distancia temporal respecto al SBI. Por el contrario, los organismos encaran entornos en los que múltiples estímulos se presentan simultáneamente y en ocasiones de manera contigua con los refuerzos. Una comida que nos enferma o que nos produce un gran placer es en sí misma un compuesto de múltiples estímulos: el plato en que se sirve, el mantel bajo el plato, cómo se ve, su aroma, la música que se está escuchando, la persona que la sirve. Más aún, cada uno de nosotros tenemos experiencias diferenciadas con cada uno de estos elementos por separado, correlacionados con otros o con el mismo reforzador. Hemos comido en ese mantel otras comidas, con platos y aromas diferentes. El modelo de Rescorla y Wagner describe el algoritmo, la regla por la cual se le asigna crédito a cada elemento de la experiencia con una comida. En ese sentido, el modelo le da respuesta a la pregunta: ¿Cómo puede un organismo extraer relaciones de "causalidad" en ésta red de diversas experiencias? En resumen, el modelo de Rescorla y Wagner captura los principios que describen la asignación de crédito a los distintos elementos de un estímulo compuesto que es seguido por un reforzador. Al mismo tiempo, el modelo especifica el efecto que juega la experiencia previa del agente con cada uno de los elementos por separado dentro de la asignación de crédito.

9.0.2 Supuestos del Modelo de Rescorla y Wagner

9.0.2.1 El supuesto de la separabilidad de los estímulos.

Los estímulos en compuesto están conformados por elementos (estímulos) separables. Desde esta perspectiva, una cara, por ejemplo, no es un estímulo integrado, sino un conjunto de elementos (ojos, boca, orejas, nariz, etc.).

9.0.2.2 El supuesto del valor predictivo de los elementos

Cada elemento de un compuesto, sea un estímulo o una respuesta, tienen un número ligado a ellos; a este número le llamamos Valor. El valor puede tomar números positivos pero también negativos. Cuando el valor es positivo predice la ocurrencia de un refuerzo, cuando es negativo predice su ausencia. Por esta razón, a dicho número también se le conoce como el valor predictivo del estímulo. El valor (V) se actualiza en cada ocasión que se presenta el estímulo o respuesta (EC) y el cambio en la magnitud del mismo depende de si el EC se presenta acompañado o no de un suceso biológicamente importante (R). La relación entre el valor y alguna medida de comportamiento es únicamente ordinal. Las diferencias en valor sólo predicen diferencias en el ordenamiento de alguna medida del comportamiento. En otras palabras, un elemento de un estímulo compuesto (por ejemplo, la nariz en los rostros) con un valor predictivo V de 1 no induce el doble de respuestas en un agente con relación a otro elemento del estímulo compuesto que tenga un valor de 0.5 (por ejemplo, el vello en los rostros): lo único que nos señalan estos valores numéricos es que el agente le está asignando mayor crédito por la ocurrencia del refuerzo a la nariz con relación al vello de los sujetos, y que el agente responderá más ante estímulos que contengan narices que ante estímulos que contengan vello (sin especificar cuantitativa ni precisamente esta diferencia de respuestas).

9.0.2.3 La regla de integración del valor de los elementos

El modelo computa por separado, para cada uno de los elementos de un compuesto, su valor predictivo V, y el valor predictivo del compuesto es la suma de los valores predictivos de cada uno de sus elementos. Si el compuesto incluye dos estímulos A y B, se computan por separado VA y VB. La fuerza de la predicción del compuesto es la suma de los Vs es, en nuestro caso:

Vtotal = VA+ VB.

9.0.2.4 La regla de la actualización del valor predictivo de los elementos.

La ecuación de Rescorla y Wagner mantiene el supuesto de que la asignación de crédito a cada uno de los elementos de un compuesto es una función de la discrepancia entre lo que se obtiene y lo que se espera obtener. La contribución de Rescorla y Wagner es suponer que lo que se espera obtener dada la presentación de un compuesto es el resultado de la suma del valor predictivo de todos los elementos presentes simultáneamente (V_total)._

Vx_t+1 = Vx_t + a(R-Vtotal_t)_

Recuerden que en nuestro protocolo, R es un valor binario que representa la presentación (R=1) o no (R=0) de un refuerzo. Como en la ecuación de Bush

y Mosteler, a es un parámetro de aprendizaje que determina la importancia del error de predicción.

La ecuación especifica la reducción del error de predicción como motor del aprendizaje y, como puede verse en el simulador, este modelo produce curvas de aprendizaje de ganancias decrecientes, en las cuales el cambio en V es cada vez más pequeño conforme el error de predicción se reduce. La parte novedosa de la ecuación consiste en tomar como predicción la suma del valor de todos los elementos presentes: lo cual nos conduce al último supuesto del modelo…

9.0.2.5 Competencia entre los elementos de un compuesto

Los elementos separados compiten entre sí por el valor predictivo del compuesto que conforman. Recordemos que el valor predictivo total de un estímulo compuesto es limitado, lo que implica que mientras mayor sea el valor de uno de los elementos, quedará menos "valor predictivo" para ser distribuido a los demás elementos del compuesto.

9.0.2.6 La ecuación de Rescorla y Wagner

Vx_{t+1} = Vx_t + alphabeta(R-Vtotal_t)

En ocasiones la ecuación de Rescorla y Wagner se representa en términos de los cambios de ensayo a ensayo.

triangle Vx =Vx_{t+1} - Vx_t

Las preguntas que emergen al considerar este modelo son: primero, ¿de qué variables depende el parámetro "a"? y segundo, ¿es "a" el único parámetro que determina la velocidad del aprendizaje? Empíricamente, podemos considerar dos variables: en primer lugar, la naturaleza del refuerzo. Por ejemplo, más y mejor comida produce un aprendizaje más rápido. Una segunda variable es la naturaleza del estímulo predictor. Un estímulo más intenso o sobresaliente produce curvas de aprendizaje más aceleradas. La importancia de este segundo elemento -la saliencia del EC- se representa en la ecuación con un parámetro adicional de aprendizaje que llamaremos beta, el cual también adquiere valores entre cero y uno y también se multiplica por el error de predicción para ponderar su importancia relativa.

triangle Vx = alpha beta (R-Vtotal_t)

Para ayudar a entender el modelo de Rescorla y Wagner en su aplicación a la vida cotidiana, consideren el siguiente escenario. Un amigo al que ustedes visitan con frecuencia consiguió un nuevo perro. A ustedes les gustaría saber si este es un perro al que se le puede acariciar sin temor a que este los muerda. El perro es un compuesto de múltiples elementos: tamaño, hocico, ojos, orejas, tipo de pelo, entre otros. Su primera respuesta ante ese nuevo perro va a ser

el resultado de la suma de los valores predictivos de los distintos elementos que lo componen, adquiridos de sus múltiples experiencias con otros perros. Por ejemplo, imaginemos que en algún momento del pasado se encontraron con un perro pequeño y chato que nunca trató de morderlos; posteriormente, cuando se encuentran con el perro de su amigo que comparte el mismo tamaño chico de aquel perro pero que tiene un hocico largo, su predicción sobre si este los morderá será la suma de lo que para ustedes predicen, por separado, su tamaño y su tipo de hocico. En este caso, el tamaño (y no el tipo de hocico) del perro de su amigo tendrá un valor predictivo en el sentido de que el animal no les morderá. Por otra parte, si el perro de su amigo intenta morderlos, el valor de los dos atributos se actualizará a través del error de predicción. Es decir, si la suma de los valores predictivos de los elementos del perro de su amigo predijo en un inicio que este no los mordería, y este efectivamente procede a morderlos, entonces habrá un error de predicción que actualizará el valor de cada uno de los elementos que conforman al perro. De esta forma, el elemento del tamaño chico del perro perderá su valor como un predictor de una "no mordida", mientras que el elemento del "hocico largo" adquirirá valor como un predictor de una "mordida".

9.0.3 Aplicación del modelo de Rescorla y Wagner al experimento de ensombrecimiento

Considere el protocolo experimental mostrado en la figura x. Hay tres grupos, para el grupo G1 en cada ensayo se presenta un compuesto de dos estímulos (tono y luz) seguidos por el acceso a alimento. Para otros dos grupos solo se les presenta el tono o la luz, cada uno seguido de comida. Vamos a suponer que el tono y la luz no son igualmente sobresalientes (es decir, tienen betas diferentes). Para el grupo G1, el error de predicción es R menos la suma del valor adquirido en cada ensayo por los elementos del compuesto, para el cual: Vtotal = Vluz + VTono. Para los otros dos grupos, el error de predicción es R menos el valor de cada elemento por separado, VT o VL. En la simulación puede verse que cuando se tiene un tono ligeramente más sobresaliente que la luz en el grupo con el estímulo compuesto, ni el tono ni la luz alcanzan valores cercanos a R. En los grupos en los cuales los dos estímulos se presentan por separado, ambos estímulos alcanzan valores que se aproximan a 1, el valor de R.

imagen

9.0.4 Aplicación del modelo de Rescorla y Wagner al experimento de bloqueo

La figura x muestra el protocolo experimental del procedimiento de bloqueo. Existen dos grupos, los cuales tienen en común el que se les presenta un compuesto de un tono y una luz, seguidos por el acceso a comida. Ambos grupos

difieren en que para uno de ellos, al cual llamaremos el grupo de bloqueo, en una primera fase se le presenta solo el tono seguido de la comida. El otro grupo, un control, no tiene esta experiencia. Para el grupo de bloqueo, que recibe en la fase 1 la experiencia con el tono seguido de la comida, al final de esa fase el valor (R - Vtono) es casi cero y el valor del tono VT es igual a R. En lenguaje menos técnico, el tono predice perfectamente la presentación de la comida.

Para este grupo en la fase en el primer ensayo de esa fase Vtotal = VT + VL = (R + 0) y consecuentemente (R - Vtotal) = (R - R +0) = 0 . Computando la actualización del valor de la luz: VL+1 = VL + a (R - Vtotal) = 0 + a (1 - 1) = 0

Vemos que no se le asigna valor al elemento luz. En resumen, cuando el elemento de un compuesto ya predice la presentación del refuerzo, el otro elemento del compuesto no adquiere valor predictivo, tal y como puede verse en el resultado de la simulación presentada en la Figura x.

imagen

9.0.4.1 Predicción contraintuitiva del modelo de Rescorla y Wagner

Cualquier versión del modelo de refuerzo predice que un refuerzo adicional debe incrementar, aunque sea por un monto muy pequeño, el valor predictivo de un estímulo. Sin embargo, veamos qué predice el modelo de Rescorla y Wagner en el siguiente protocolo. A un grupo lo exponemos a tres fases de entrenamiento. En la primera fase, un tono es seguido de comida durante 60 ensayos. En una segunda fase, una luz es el estímulo condicionado y es seguida de comida durante otros 60 ensayos. En la tercera fase, la final, a los sujetos experimentales se les presenta el compuesto tono-luz, seguido de comida durante 60 ensayos. Al inicio de la tercera fase, VL = R; VT = R y Vtotal = 2R, de tal forma que el error de predicción para ambos estímulos, será R - 2R, por lo que Vt+1 será un número negativo y veremos un decremento en el valor predictivo para los dos estímulos. Interesantemente, se ha encontrado evidencia empírica que respalda esta predicción contraintuitiva del modelo. La siguiente figura muestra el resultado de la simulación.

imagen

9.0.4.2 Aplicación del modelo de Rescorla y Wagner a estudios de protocolos de correlaciones

Un reto importante para la ecuación de Rescorla y Wagner es dar cuenta de los resultados de los experimentos de Rescorla en los que se manipula la relación de contingencia: esto es, experimentos en los que se manipula la probabilidad de la presentación del refuerzo, dada la presencia o ausencia del EC. Recuerden que en esos experimentos, se encontró que manteniendo constante la probabilidad

de refuerzo en la presencia del estímulo condicionado, el crédito que se le asigna depende de la probabilidad de refuerzo en su ausencia. Sin embargo, de acuerdo a una interpretación literal de la ecuación de Rescorla y Wagner, el error de predicción para el estímulo condicionado es independiente de la aparición o no aparición del refuerzo en la ausencia del EC. La solución propuesta por Rescorla y Wagner para que su modelo de cuenta de estos hechos empíricos es considerar al contexto en el que se presenta el EC como un estímulo más. El contexto es el interior del espacio experimental e incluye, entre otros elementos, la iluminación, el olor y la textura del espacio. De esta forma, el protocolo de los experimentos de Rescorla incluye dos estímulos: el compuesto del estímulo condicionado junto con el contexto; y un segundo estímulo, el contexto solo. En el caso del procedimiento con igual probabilidad de refuerzo en la presencia y la ausencia del EC, la ecuación de Rescorla y Wagner interpreta el experimento como uno de bloqueo en el que el contexto X es el mejor predictor del refuerzo y termina bloqueando la asignación de crédito al estímulo condicionado.

Una forma de evaluar su comprensión del modelo de Rescorla y Wagner es considerar cuál sería su predicción para un experimento con protocolo no correlacionado (sin correlación entre el EC y R), en el cual los refuerzos que se presentan durante el intervalo entre ensayos son señalados con un tercer estímulo diferente al estímulo condicionado. ¿Qué cambios se pueden esperar en la asignación de crédito el estímulo condicionado?

9.0.5 El modelo de Rescorla y Wagner e inhibición condicionada

Hasta este punto, hemos argumentado que de acuerdo al modelo de Rescorla y Wagner, tanto estímulos como respuestas adquieren un valor que les permite predecir la presencia de un refuerzo, pero siguiendo este modelo ¿pueden los estímulos/respuestas predecir la ausencia de un refuerzo? En este apartado le daremos respuesta a esa pregunta.

Poder predecir la *no* ocurrencia de ciertos refuerzos, tiene importantes ventajas competitivas para el organismo, en particular, le permite acomodar su distribución de comportamientos de una mejor manera. La señal de que un depredador no va a aparecer, le permite a la potencial presa buscar su alimento sin interrupciones; de igual manera, la señal que predice que no habrá comida, le permite al organismo reorientar su comportamiento hacia la búsqueda de otros refuerzos. Al estudio de este fenómeno se le conoce como *inhibición condicionada*.

El estudio de la inhibición condicionada tardó décadas en despegar por dos razones. La primera está relacionada con la estructura del modelo original de aprendizaje por refuerzo, que no permite valores negativos para el valor de un estímulo. Si la mayor parte del flujo de estímulos y respuestas no van seguidos de un refuerzo, ¿que se aprende acerca de estos eventos? Imaginen que se encuentran con una persona paseando a un perro que los ignora completamente.

Consideremos qué predice el modelo de refuerzo sobre lo que ustedes aprenderán acerca de esa persona. En este episodio, el perro no era un suceso biológicamente importante -ni les gruñó, ni les movió la cola- consecuentemente R es igual a cero. Adicionalmente, la persona era un desconocido que no predice nada, su V es por lo tanto igual a cero. De acuerdo al modelo de refuerzo, el cambio en el valor predictivo de la persona (Vx) es una función del error de predicción (R - Vx), en este caso (0 - 0) y por lo tanto no habría tampoco ningún cambio en Vx. En otras palabras, si dado un estímulo, nada se espera y nada se obtiene, ese estímulo no predice nada.

A diferencia del modelo de refuerzo tradicional, el modelo de Rescorla y Wagner permite que un estímulo tenga un valor negativo y sea un predictor de la ausencia de un reforzador. Regresemos a nuestro ejemplo de una persona paseando a un perro, excepto que esta vez, imaginemos que el perro les gruñe de forma amenazante. Después de muchos encuentros similares, la persona paseando al perro se convierte en el predictor de un perro agresivo. En un siguiente encuentro, la persona que pasea al perro va acompañada de su pareja y el perro esta vez no les gruñe, *generando un error de predicción* con valor negativo. Después de muchos encuentros de este tipo, la pareja de la persona que pasea al perro se convierte en un inhibidor condicionado, el cual predice la no ocurrencia del gruñido del perro. Recordando que en la ausencia de un refuerzo, R es igual a cero, el error de predicción es negativo sólo si el estímulo neutro aparece en compuesto con un estímulo con valor positivo. De esa forma V total > 0 y el error de predicción (R - V total) < 0.

En conclusión, de acuerdo a Rescorla y Wagner, un estímulo/ respuesta se convierte en un inhibidor condicionado, solo si hay un error de predicción negativo como resultado de presentarlo en compuesto con un predictor de refuerzo.

La segunda razón que dificulta el estudio de la inhibición condicionada es la dificultad para distinguir empíricamente entre un estímulo neutro -es decir, uno que no predice nada- y un estímulo que predice la ausencia de algo. En este tema, la contribución de Rescorla es también un punto de partida. En 1969, propuso dos protocolos necesarios para argumentar y sostener que un estímulo era un inhibidor condicionado.

En un primer protocolo, conocido como de *sumación*, se compara, por un lado, la respuesta a un estímulo condicionado A con valor positivo presentado individualmente; con, por otra parte, la respuesta ante un compuesto del estímulo A acompañado de un estímulo X. Nuestro objetivo es determinar si X es un inhibidor condicionado. Si la respuesta al compuesto AX es menor u opuesta a la respuesta observada ante el estímulo A presentado individualmente, podríamos concluir que el estímulo X es un inhibidor condicionado. Regresando a nuestro ejemplo, podemos argumentar que la pareja del paseador de perro es un inhibidor condicionado, si el perro no gruñe cuando la pareja acompaña al paseador y sí gruñe cuando este va acompañado únicamente de su paseador. Sin embargo, Rescorla señala que estos resultados tienen una segunda interpretación: es posible que la atención dirigida al estímulo X (la pareja) reduzca

la atención dirigida al estímulo A (el paseador) resultando en la menor respuesta al perro. En resumen, X no sería un predictor de la ausencia de gruñido (no sería un inhibidor condicionado), simplemente, X contribuiría a que se ignore a A. Las siguientes dos figuras muestran el protocolo de sumación y el resultado de una simulación.

IMAGENES

De acuerdo a Rescorla, para descartar la interpretación alternativa del protocolo de sumación en términos de atención se requiere de una prueba adicional. A esta se le conoce como *prueba de retardo* y consiste en comparar la curva de adquisición de valor predictivo de un estímulo neutral A presentado individualmente, con la presentación -también individual- de un estímulo X que se entrenó como un inhibitorio y que tiene un valor negativo. Si el aprendizaje es más lento para el segundo estímulo X, podríamos concluir que este se trata de un estímulo inhibitorio. La figura x muestra los resultados de la simulación con este protocolo.

IMAGEN

Sin embargo, otra posible explicación de los resultados de la prueba de retardo hace también referencia a procesos de atención. Es posible que por el entrenamiento previo, el estímulo X deje de activar los procesos de atención y que por lo tanto la demora en el aprendizaje de su valor predictivo se deba a la falta de atención que este recibe en comparación con la atención que recibe el estímulo novedoso neutral. Sin embargo, noten que en el protocolo de sumación, la explicación alternativa para dar cuenta de una menor respuesta del organismo ante el estímulo compuesto (X+EC) era una mayor atención otorgada al estímulo X, el cual desviaba la atención del estímulo EC; mientras tanto, en el protocolo de retardo, la explicación alternativa para dar cuenta de la menor respuesta ante el estímulo X es una menor atención asignada al estímulo X debido a la familiaridad con este estímulo. Por lo tanto, Rescorla propone que para concluir que un estímulo/respuesta es un inhibidor, este debe pasar tanto la prueba de sumación como la de retardo. No resulta plausible que un estimulo reciba menos atención en una circunstancia y el mismo reciba mayor atención en la otra circunstancia: por lo cual, si el estímulo actúa como inhibidor en ambas circunstancias, esto significa que su efecto no se debe a meros procesos atencionales de novedad/habituación, sino que efectivamente, el organismo considera a este estímulo como un predictor de la ausencia de un SBI. En otras palabras, dado que la variable de atención tiene efectos contrarios en las dos pruebas: al encontrar un estímulo que pasa ambas pruebas, se eliminan las explicaciones alternativas de más y de menor atención, y se puede considerar a este estímulo como un inhibidor condicionado.

9.0.6 Algunos problemas con el modelo de Rescorla y Wagner

Dentro de la Psicología, pocos modelos han sido tan exitosos como el de Rescorla y Wagner en dar cuenta de una amplia gama de resultados, abrir nuevas rutas de investigación y capturar formalmente explicaciones alternativas al papel de la contigüidad en la asignación de crédito. Sin embargo, como ocurre con cualquier otro modelo, hay un número de sus predicciones que no tienen sustento empírico. Estas fallas han dado lugar a extensiones de modelos y a modelos alternativos que veremos en otro capítulo. A continuación presentamos dos de las predicciones erróneas. Seleccionamos estas por su fácil comprensión y por ser las que han dado lugar a modelos alternativos.

9.0.6.1 Supuesto de que la extinción reduce el valor de un estímulo/respuesta a cero

En extinción, el refuerzo que previamente seguía a un estímulo o respuesta deja de presentarse. En este caso R cambia de un valor de 1 a 0. El modelo asume que no habrá un error de predicción cuando el valor del estímulo sea igual a R, esto es, cuando el valor del estímulo sea también cero. Consecuentemente, para Rescorla y Wagner, el impacto de un estímulo que fue extinguido, debe ser el mismo que el de un estímulo neutral dado que para ambos estímulos $V = 0$.

Sin embargo, hay una multitud de reportes que señalan que el mero paso del tiempo produce una recuperación espontánea del efecto que originalmente tenía un estímulo que ha atravesado un proceso de extinción. Adicionalmente, se ha encontrado que estímulos previamente extinguidos adquieren valor predictivo más rápido que estímulos neutrales, a pesar de que se supone que ambos inician con un valor igual a cero. Esta literatura y los modelos para dar cuenta de ella la presentamos en el capítulo sobre extinción.

9.0.7 Inhibición latente

Consideremos el siguiente protocolo experimental. En una primera fase, a un grupo se le presenta, durante 60 ensayos, un estímulo neutral que no es seguido por un refuerzo; y en una segunda fase, la presentación de este mismo estímulo sí va seguida de un refuerzo. A un segundo grupo, solo se les presenta la segunda fase, sin darle la experiencia con el estímulo solo. A este protocolo se le conoce como inhibición latente. De acuerdo a Rescorla y Wagner, el valor de los dos estímulos debería ser el mismo. Sin embargo, una enorme literatura reporta que la preexposición a un estímulo sin refuerzo demora la subsecuente adquisición de valor cuando ese estímulo va acompañado de un refuerzo.

El fenómeno de la inhibición latente ha generado una importante alternativa teórica al modelo de Rescorla y Wagner, la cual revisaremos en otro capítulo y

que pone el énfasis en la *atención* asignada a un estímulo cuando este es seguido de un refuerzo inesperado. Inhibición latente sería el resultado de la falta de atención que se le asigna a un estímulo que fue presentado por muchos ensayos sin ser seguido por ningún refuerzo, y que por lo tanto, desde la perspectiva del organismo, no predice que algo importante ocurrirá.

Chapter 10

Acción Como Elección

Todas las variantes de la ley del efecto nos dicen que las respuestas que son seguidas por un refuerzo incrementan en frecuencia. Si esta fuera la única información que nos proporciona este concepto, difícilmente le asignaríamos el estatus de una una ley y su utilidad para entender el comportamiento sería muy limitada. En un escenario aplicado, para la madre que quiere modificar la conducta de un hijo, la sola recomendación de reforzar la conducta que desea incrementar no resulta totalmente satisfactoria. La madre quisiera saber si es necesario reforzar cada instancia de la respuesta o solo algunas de ellas, quisiera saber si el refuerzo 20 tiene el mismo impacto que el refuerzo 10 y, finalmente, quisiera saber si el seguimiento de distintas reglas para la entrega del refuerzo marca una diferencia.

En las notas sobre programas de refuerzo, le dimos ya una respuesta a la última pregunta: presentamos el comportamiento característico asociado con distintas reglas de entrega de refuerzo y, en particular, resaltamos que en programas de razón variable los organismos responden a una tasa más alta que en los programas de intervalo variable. En el escenario aplicado, la madre preguntaría si para conseguir que su hijo emita la respuesta deseada a cierta tasa, el valor de la razón debe ser de 5, 10, 50, 200 o 500 respuestas por refuerzo. La madre también quisiera saber qué diferencia hace para el comportamiento observado que la oportunidad de obtener un refuerzo sea de 1 en cada 5, 10, 20, 100, 200 o 500 minutos.

Por lo tanto, lo que le hace falta a la ley del efecto con la que iniciamos esta nota es especificar la función que describe la relación entre la tasa de respuesta de un organismo y una medida del refuerzo. Para propósitos de estas notas, la medida del refuerzo será la tasa de ocurrencia del reforzador:

$$R_i = f(r_i)$$

Un primer paso para especificar la función aludida es encontrar la relación empírica entre la tasa de respuesta de los organismos y distintos valores dentro de programas de intervalo variable y de razón variable.

10.1 Funciones de Respuesta para programas de intervalo variable

Catania y Reynolds (1968) publicaron el primer estudio sistemático en el que se estudió la relación, "en equilibrio", entre tasas de respuesta y tasas de refuerzo que resultan de variar los valores del programa de intervalo. Al hablar de "equilibrio", se hace referencia a un protocolo en el que se entrena al animal con un valor de un programa IV y se ejecutan el número de sesiones necesarias para alcanzar un comportamiento estable; finalmente, los datos que se analizan son los correspondientes a los últimos días de exposición. El procedimiento se repite para cada uno de los programas IV estudiados. De esta manera, se determina la relación entre la tasa de refuerzo de distintos intervalos (valor de entrada de la función) y la tasa de respuesta de los organismos (valor de salida de la función).

imagen

En la siguiente figura se muestran los resultados del estudio de Catania y Reynolds. Puede verse que hay una relación de ganancias decrecientes entre la tasa de respuesta y la tasa de refuerzo que produce. La función crece rápidamente conforme incrementa la tasa de refuerzo (es decir, conforme aumenta el número de refuerzos por unidad de tiempo), hasta alcanzar un punto después del cual, incrementos subsecuentes en la tasa de refuerzo tienen un efecto cada vez menor sobre la tasa de respuesta del organismo. Esta función se encontró posteriormente en docenas de estudios (de Villiers y Herrnstein).

10.1.1 Relación Entre Tasas Absolutas y Relativas de Respuesta

Para Herrnstein (1970) fue claro que cualquier función que relacione la tasa de respuesta (como valor de salida) a la tasa de refuerzo (como valor de entrada) debe satisfacer dos condiciones: * primero, dar cuenta de la relación obtenida en los programas de intervalo variable cuando hay una sola opción de respuesta disponible para el organismo y * segundo, ser consistente con la ley de igualación, cuando hay dos o más opciones de respuesta para el organismo.

Recordemos que para los modelos de refuerzo, los animales no llevan un registro de la frecuencia relativa con la que deben emitir cada respuesta, en lugar de ello, estos modelos postulan que para el organismo cada respuesta adquiere un valor por separado, los cuales subyacen a los patrones de respuesta observados. El valor de cada respuesta varía en función del refuerzo que ésta produce. Por

otro lado, la tasa relativa con la que el organismo emite cada respuesta es el resultado de una comparación entre el valor adquirido por cada una de las distintas respuestas. El objetivo, por lo tanto, es encontrar la función que describe la relación entre los tres valores: el valor adquirido por cada respuesta; la tasa con la que el organismo emite cada respuesta; y las tasas de refuerzo que resultan de cada una de las tasas de respuesta:

$$R_i = f(r_i)$$

Y usando la función *f*, derivar la relación entre las tasas relativas de respuesta y las tasas relativas de refuerzo. La función f la vamos a evaluar por su éxito para dar cuenta de las tasas relativas de respuesta (siguiendo el patrón de la ley de igualación) observadas en programas concurrentes IV - IV. Para ello, debemos tener presente que la programación de los refuerzos en lo programas concurrentes puede hacerse de dos formas:

- La primera es manteniendo constante la tasa de refuerzo a lo largo del experimento, pero variando la proporción asignada a cada respuesta en diferentes condiciones experimentales. Por ejemplo, se pueden programar 60 refuerzos por hora, al mismo tiempo que se establecen diferentes condiciones experimentales en las que se generan las siguientes distribuciones de refuerzos para las dos respuestas posibles: 50-10, 40-20, 30-30, 20-40 y 10-50.
- La segunda es programando una tasa de refuerzo fija para una de las respuestas a lo largo del experimento, a la vez que se varía la tasa de refuerzo para la segunda respuesta en diferentes condiciones experimentales.

La siguiente figura muestra los resultados de un experimento que compara los resultados de las dos formas de estructurar los programas concurrentes:

imagen

La figura del panel izquierdo muestra la tasa de respuesta de las dos opciones bajo la condición donde la tasa de refuerzo total es constante. La figura del panel derecho muestra la tasa de respuesta de las dos opciones bajo la condición en la cual la tasa de refuerzo para una respuesta (símbolos negros) es constante, mientras que la tasa de refuerzo para la otra respuesta es variable (círculos abiertos). Podemos ver que en la primera condición, la tasa de respuesta está linealmente relacionada a su tasa de refuerzo. A mayor tasa de refuerzo, mayor tasa de respuesta. Mientras tanto, en la segunda condición, la tasa de respuesta a la opción con refuerzo fijo disminuye en función del aumento en el refuerzo obtenido en la otra opción de respuesta.

En otras palabras, la diferencia sútil pero crucial entre las dos condiciones reside en cómo la tasa de respuesta de una opción varía con relación a la tasa de respuesta de la otra opción dentro de una misma condición.

En la primera condición (refuerzo total constante), si la tasa de refuerzo de una opción aumenta, la tasa de refuerzo de la otra opción decrementa proporcionalmente (dado que el total se encuentra fijo). Esto genera una compensación perfectamente lineal a nivel de las tasas de refuerzo: y las tasas de respuesta reflejan esta relación lineal porque esencialmente ambas opciones compiten por un pool de respuestas fijo.

En la segunda condición (una opción con refuerzo fijo, la otra con refuerzo variable), el aumento a la tasa de refuerzo para la opción variable disminuye la tasa de respuesta para la opción fija. A pesar de que la tasa de refuerzo de la opción fija no ha cambiado en absoluto, las respuestas a ella disminuyen de todas formas cuando la segunda opción se vuelve más recompensante. Esto demuestra que el organismo no mantiene meramente un registro de las tasas de refuerzo absolutas, sino que ejecuta una computación más compleja sobre el valor relativo de cada opción dentro del contexto más amplio.

10.1.2 Posibles Funciones de Refuerzo

La primera función que podemos considerar tiene su origen en una propuesta de Skinner. De acuerdo a esta, la tasa de respuesta es directamente proporcional a la tasa de refuerzo.

$$R_i = kr_i$$

La ecuación describe una línea recta, donde el parámetro **k** es su pendiente y representa la traducción de un refuerzo en un incremento fijo en la tasa de respuesta. (Puede trabajar con el correspondiente simulador).

Para determinar si esta primera función es consistente con la ley de igualación, insertamos la función en la computación de tasas relativas de respuesta. Asumiendo que el refuerzo es el mismo para las dos opciones, el parámetro **k** se cancela y obtenemos la ecuación de igualación.

$$\frac{R_1}{(R_1 + R_2)} = \frac{kr_1}{(kr_1 + kr_2)}$$

La función que postula proporcionalidad entre respuestas y refuerzos es consistente con los resultados presentados en la Fig x. Sin embargo, la función

$$R_i = kr_i$$

propone que la tasa de una respuesta depende únicamente del refuerzo que ésta produce. Por lo tanto, la tasa de una respuesta no debería de cambiar cuando su refuerzo es constante y tan solo se manipula otra fuente de refuerzo. Sin embargo, en el panel derecho de la figura x observamos que la tasa de

respuesta con el refuerzo constante decrece conforme incrementa el refuerzo para la respuesta alternativa: fenómeno que se conoce como **contraste conductual**.

La siguiente es una segunda función de respuesta, consistente con el muy replicado resultado de contraste conductual:

$$R_1 = \frac{kr_1}{(r_1 + r_2)}$$

La función nos dice que la tasa de una respuesta es una función de la tasa relativa de refuerzo que recibe esta respuesta. Como puede verse en el simulador, si la tasa de refuerzo r_2 para la otra opción es constante, la forma de la función de la tasa de respuesta R_1 es de ganancias decrecientes, igual a los datos empíricos reportados en la figura x. Por otra parte, si r_1 es constante, conforme incrementa el valor de r_2, la tasa de de respuesta R_1 decrementa, reproduciendo el patrón de contraste conductual.

Esta función es consistente con el resultado de la ley de igualación. Sustituyendo y cancelando el parámetro k y los denominadores,

$$\frac{R_1}{R_1 + P_2} = \frac{\frac{kr_1}{r_1+r_2}}{\frac{kr_1}{r_1+r_2} + \frac{kr_2}{r_1+r_2}} = \frac{kr_1}{kr_1 + kr_2} = \frac{r_1}{r_1 + r_2} \tag{10.1}$$

Esta segunda función da cuenta de los resultados obtenidos en programas concurrentes. Sin embargo, veamos qué predice la misma función para experimentos en los que solo se refuerza una de las respuestas, como es el caso de los datos del experimento de Catania y Reynolds. En la ecuación:

$$R_1 = \frac{kr_1}{(r_1 + r_2)}$$

el valor de r_2 es cero, pues solo hay una respuesta (r_1) y la ecuación se reduce a

$$R_1 = \frac{kr_1}{r_1}$$

Por lo tanto, terminamos con la función:

$$R_1 = k$$

la cual nos dice que la tasa de respuesta es una constante y que por ende, la respuesta es insensible a la tasa de refuerzo. Sin embargo, esta conclusión matemática no corresponde a los resultados reportados en la revisión de Villiers y Herrnstein sobre múltiples experimentos similares a los de Catania y Reynolds.

En ella, se observa que para protocolos con una única opción de respuesta, dos cosas son ciertas: la tasa de respuesta NO crece linealmente en función de la tasa de refuerzo, contrario a lo que sugiere la primera función de Skinner; y, de igual manera, la tasa de respuesta TAMPOCO es una constante insensible a variaciones en la tasa de refuerzo, contrario a lo que sugiere la segunda función de refuerzo.

10.1.3 La ley del Efecto Relativa

En un artículo publicado en 1970, Herrnstein propuso una versión de la ley del efecto que simultáneamente daba cuenta de dos de las regularidades empíricas más robustas: la igualación en programas concurrentes y la función de ganancias decrecientes entre tasa de respuesta y la tasa de refuerzo. Propuso un planteamiento que en su momento fue una gran salto:

Todo comportamiento es una instancia de una elección y el papel del refuerzo es redistribuir el comportamiento.

La propuesta descansa en tres supuestos:

1. Los organismo están en constante actuar y no existen vacíos conductuales. En toda situación experimental hay al menos dos respuestas, una, R_1, que medimos directamente (como picar una tecla, por ejemplo) y un conjunto de respuestas que no medimos directamente (como dar vueltas, aletear, picar el piso, etcétera) y que que Herrnstein llamó R_o. La suma de estas dos respuestas componen la totalidad del comportamiento k.

$$R_1 + Ro = k$$

2. En adición al refuerzo programado en una situación experimental, siempre hay otros refuerzos disponibles para el organismo, llamados *ro*.

3. Los organismos igualan la frecuencia relativa de la respuesta registrada y de todas las otras respuestas con la tasa relativa de refuerzo de todas las respuestas en una situación experimental.

$$\frac{R_1}{R_1 + R_0} = \frac{r_1}{(r_1 + r_o)}$$

$$R_1 + R_0 = k$$

$$\frac{R_1}{k} = \frac{r_1}{(r_1 + r_o)}$$

$$R_1 = \frac{kr_1}{(r_1 + r_o)}$$

Recordemos que k es un parámetro que se estima estadísticamente y representa la suma de todos los comportamientos medida en unidades R_1: frecuentemente respuestas por minuto. El valor de k depende de la topografía (la forma específica) de la respuesta R_1 que se mide, así como del tiempo que toma ejecutarla. Picar una tecla es mucho más fácil para una paloma de lo que es apretar una palanca para una rata: como consecuencia, el valor de k es mayor cuando se mide el comportamiento de picar una tecla que cuando se mide el comportamiento de apretar una palanca.

r_o es un segundo parámetro que se estima estadísticamente y que representa la tasa de otros refuerzos no controlados en el experimento. r_o se mide en las mismas unidades que r_1: esto es, el número de refuerzos recibidos por unidad de tiempo.

En el simulador, ustedes pueden ver cómo varía la forma de la función que relaciona la tasa de respuesta a la tasa de refuerzo cuando se varía el valor de r_o dentro de programas de intervalo variable. En general, estos modelos exponen cómo el número de refuerzos por unidad de tiempo generados por una respuesta (dentro de un programa de intervalo variable X) influye sobre la tasa con la que el organismo emite esa respuesta; al mismo tiempo, la tasa de refuerzo que generan las respuestas no controladas (R_o) influye sobre la tasa relativa de la respuesta evaluada (R_1).

- La función R_1 es de *ganancias decrecientes*, esto es, el impacto de un refuerzo adicional es **mayor** cuando r_1 es pequeño y va **decreciendo** conforme r_1 incrementa. Ver figura.
- La función nos muestra que el impacto de un refuerzo contingente sobre una respuesta *depende del contexto de refuerzo.* A medida que aumenta r_o, el impacto de incrementar r_1 es menor. Con valores de r_o muy pequeños, un contexto de refuerzo muy pobre, la tasa de respuesta es muy sensible a cambios en la tasa de refuerzo r_1 y justo lo opuesto ocurre con un contexto de refuerzo muy rico, donde r_o es muy grande. Para una persona en pobreza extrema, pequeños cambios en el monto de su pensión tienen gran impacto en su comportamiento, pero no así para una persona en un entorno de riqueza. Ver figura.

10.1.4 Impacto sobre la modificación de la conducta

La ley del efecto relativo ha tenido un gran impacto sobre la práctica de la modificación de la conducta. Considere un caso en el que se pretende reducir un comportamiento indeseable en un niño, como podría ser un comportamiento

agresivo. De acuerdo a la versión original de la ley del efecto, una posibilidad es eliminar el refuerzo para ese comportamiento. Desafortunadamente, el refuerzo para la conducta agresiva del agresor puede ser la reacción de sumisión y respeto que este comportamiento induce en el niño agredido, sobre la cual no es fácil tener un control. Otra posibilidad es castigar al niño agresivo, arriesgando el surgimiento de otras respuestas indeseables. El modelo de Herrnstein proporciona una alternativa viable y exitosa: esta consiste en seleccionar y reforzar un comportamiento incompatible con el agresivo. Por ejemplo, se le pide a la maestra reforzar y jugar un juego de mesa con el niño. Para involucrarse plenamente en el juego de mesa, el niño tiene que sacrificar otros de sus comportamientos, entre ellos, sus comportamientos agresivos.

La estrategia puede extenderse a problemas tan severos como el consumo excesivo de bebidas alcohólicas. Una de las consecuencias del alcoholismo es el creciente aislamiento de la persona, con la asociada reducción en el contexto de refuerzo social. Esto genera un círculo vicioso, puesto que la reducción en el contexto de refuerzo social aumenta el impacto del refuerzo asociado con la bebida alcohólica. De acuerdo al modelo de Herrnstein, una estrategia exitosa sería recuperar la vida social de la persona, incrementando así los refuerzos que conforman su contexto cotidiano.

Similarmente, si se desea incrementar un comportamiento, la estrategia sugerida por el modelo de Herrnstein consiste en reducir las otras fuentes de refuerzo disponibles para la persona. El impacto de reforzar la conducta de estudio de un niño es mayor si se eliminan las opciones de la televisión y de un celular.

En resumen, la ley del efecto relativa, al dirigir el estudio del comportamiento al estudio de la elección, proporciona una novedosa alternativa a las prácticas tradicionales de la modificación de la conducta. En lugar de simplemente eliminar o castigar conductas indeseables, se busca promover alternativas deseables y modificar el contexto de refuerzo para influir en las elecciones del individuo.

10.1.5 Evaluación

Empíricamente, la ley del efecto relativo de Herrnstein es enormemente exitosa en su descripción de la relación entre tasa de respuesta y la tasa de refuerzo dentro de programas de intervalo variable (de Villiers y Herrnstein). Sin embargo, la ecuación es algo más que un ejercicio para el ajuste de datos empíricos, esta es el resultado de un conjunto de supuestos teóricos que se reflejan en la interpretación de los dos parámetros de la ecuación k y r_o.

Las siguientes son algunas de las predicciones teóricas del modelo:

- La suma del total de respuestas, capturada por k, debe ser constante e independiente de variables que afecten el valor del refuerzo contingente y el valor de r_o, tales como manipulaciones motivacionales y la calidad del refuerzo contingente. En otras palabras, cuando se dan cambios en el

valor de los refuerzos, lo que puede cambiar es la frecuencia relativa de las distintas respuestas, pero el número total de respuestas potencialmente realizables por el organismo debe permanecer constante.
- Las manipulaciones en el tipo y la magnitud del refuerzo contingente deben ser capturadas por cambios en r_o.
- Agregar otra fuente de refuerzo debe reducir el valor del parámetro r_o. ### Quedamos de elaborar más/ejemplificar estos últimos puntos..

La evidencia acerca de los supuestos teóricos de la ley del efecto relativo no son consistentemente favorables (Dallery y Soto, 2004), en particular, en varios experimentos se reportan cambios en el parámetro k cuando se dan cambios en la magnitud de refuerzo: es decir que no sólamente se redistribuye el número de respuestas asignado a las distintas opciones de comportamiento, sino que en efecto, el número total de respuestas que el organismo emite dentro de una cierta duración temporal incrementa. Estos resultados, acompañados del hecho de que la ecuación no se ha aplicado a los resultados obtenidos en programas de razón variable, han llevado a la derivación de la ecuación de ganancias decrecientes a partir de otros supuestos, tema que abordaremos en las siguientes notas. Podemos concluir que, sin duda, la ley del efecto relativo de Herrnstein no solo generó una gran cantidad de evidencia empírica, sino que adicionalmente brindó las bases para la gran mayoría de los modelos de acción. Estas bases serán desarrolladas a detalle posteriormente, pero pueden resumirse en la siguiente lista:

1. El estudio de la acción es el estudio de la elección.
2. La acción observada se mide por su tasa de ocurrencia y refleja la distribución total del comportamiento posible.
3. El efecto del refuerzo es cambiar la distribución del comportamiento.
4. El efecto del refuerzo depende del contexto de otros refuerzos presentes, incluyendo los refuerzos no medidos directamente.
5. La regla que gobierna la distribución del comportamiento es la igualación de la frecuencia relativa de la respuesta a la tasa relativa de refuerzo con la que ésta se encuentra asociada.

Chapter 11

Elección Recurrente: Igualación

Consideren las siguientes situaciones: a un agente se le presentan dos bolsas, y se le **informa** que una contiene $100,000 y la otra $100 a la vez que se le pide que opte por una de ellas; en un segundo escenario, al agente se le lleva a un restaurante que no volverá a visitar y se le pide elegir entre uno u otro platillo. En ambos ejemplos, la elección puede hacerse exclusivamente a partir del valor de las opciones en el momento de la elección, y anteriormente hemos visto que el agente selecciona la opción con mayor valor en el momento de decisión. Estas situaciones de elección son instancias de sistemas abiertos sin retroalimentación y siguiendo a Gallistel (fecha), les llamamos protocolos de *optar*. En las últimas décadas, la mayoría de la investigación psicológica con participantes humanos ha utilizado este tipo de protocolo.

Consideren ahora un experimento similar: en lugar de informar al agente acerca del contenido de las bolsas en un inicio, a este se le presentan las opciones como bolsas cerradas, y a través de múltiples iteraciones, el agente tiene que explorar y **aprender** acerca de los montos de dinero de las distintas bolsas o de la calidad de las distintas opciones ofrecidas por el restaurante. La presentación recurrente de las oportunidades de elección crea nuevos problemas de adaptación para el agente. Como vimos en las notas anteriores, el agente debe resolver el dilema de exploración - explotación y debe determinar si las consecuencias de las opciones cambian como una función de las elecciones y el paso del tiempo, o si estas son fijas e independientes de las elecciones y del tiempo. En un ejemplo no ya de una elección entre varios platillos de un restaurante sino de una elección entre acudir a uno de varios restaurantes, el agente debe visitar los distintos locales un buen número de veces antes de tomar la decisión de pasar a explotar uno. Sin embargo, aún tras haber estimado la opción con mayor valor y haber comenzado a explotar una opción elegida, el agente debe mantener

cierta flexibilidad para modificar su elección, en aras de poder determinar si la calidad de los restaurantes varía aleatoriamente con el paso del tiempo (debido a factores como los cambios de cocinero, por ejemplo). En estos casos, la regla de elegir la opción que instantáneamente tiene más valor no es la que a largo plazo proporciona la mayor cantidad de refuerzos. En los experimentos con estos protocolos observamos una regla de respuesta probabilística.

11.0.1 Elección Recurrente

Fuera del laboratorio, lo común para los humanos y para muchas otras especies son situaciones en las cuales los individuos pueden elegir entre dos o más acciones o parcelas de forma repetida y continua, y en las cuales las elecciones alteran las opciones futuras de refuerzo. Estos problemas de adaptación son ejemplos de sistemas de retroalimentación cerrados y, siguiendo a Gallistel, les llamamos *problemas de asignación* de respuestas, tiempo o esfuerzo. Un estudiante a lo largo del día va asignando su tiempo a diferentes actividades: desayunar, transportarse, pasar tiempo en el salón de clases, estudiar en la biblioteca, conversar con amistades, ver a su pareja, ejercitarse. Al final del día, habrá una distribución de tiempos asignados a las diferentes actividades. Resulta importante considerar que cada una de estas actividades es en sí otro espacio de posibles acciones a las que se les puede dedicar tiempo. Por ejemplo, cuando están en el salón de clases, pueden atender lo que presenta el profesor o pueden ver las noticias en su celular, mandar un WA o fantasear. De esta forma, podemos estudiar problemas de asignación a diferentes escalas temporales, pero todas bajo el mismo esquema.

Un ejemplo adicional lo proporciona el forrajeo de una abeja que enfrenta dos diferentes parcelas con flores con polen. En este entorno, mientras más tiempo pasa la abeja visitando las flores de una de las parcelas, la disponibilidad del polen dentro de la misma va disminuyendo; al mismo tiempo, las flores en la otra parcela siguen llenas de polen. La abeja enfrenta dos problemas de adaptación: el primero es decidir cuánto tiempo agregado dedicarle a cada una de las dos parcelas como una función de la distribución de flores con alimento en cada una de las parcelas. Normalmente, en el contexto de asignación, la regla es distribuir el comportamiento a lo largo del tiempo de una forma que produzca la mayor ganancia posible. El segundo problema de adaptación es la decisión de cuándo salirse de una de las parcelas para visitar la otra. En estas notas, nos centraremos en el estudio del primer problema de adaptación: la distribución de respuestas y tiempos.

La forma más sencilla de estudiar experimentalmente protocolos de elección recurrente se desarrolló en el laboratorio de Skinner y se conoce como *programas de refuerzo concurrentes*. El protocolo consiste en presentarle a un organismo dos o más opciones de respuesta -teclas iluminadas en el caso de las palomas- que se encuentran disponibles todo el tiempo y que siguen programas individuales e independientes de refuerzo. (Figura). En estas notas, revisaremos los resultados

obtenidos en el estado de equilibrio, una vez que los agentes han aprendido acerca de las consecuencias de cada opción de respuesta. La variable que se estudia es la distribución de respuestas o tiempos asignados.

$$\frac{R_1}{(R_1 + R_2)}$$

imagen

En los estudios que reportamos, el animal es expuesto a un par de programas de refuerzo dentro de sesiones diarias hasta que la distribución de respuestas a las distintas opciones disponibles se vuelve estable y deja de cambiar día con día. Esto toma entre 30 y 45 días. Esta rutina se repite para todos los pares de programas que se están estudiando. De cada par, se usan para el análisis los últimos cinco días en los que las elecciones son estables.

11.0.2 La Ley de Igualación

En 1961, Richard Herrnstein (ver foto)

imagen

reportó los resultados del primer estudio con programas concurrentes, en el que la respuesta de picar una de dos teclas era reforzada de acuerdo a un programa de intervalo variable. Recuerden que en estos programas el refuerzo se presenta tras la primera respuesta después de que haya transcurrido un tiempo aleatorio desde el último refuerzo. Un detalle muy importante que debe tenerse presente bajo este tipo de programa es que una vez que un refuerzo está disponible, la oportunidad de obtenerlo se retiene hasta que el animal responde a esa opción.

Herrnstein encontró que en éstos programas la tasa relativa de respuestas (su proporción) iguala la tasa relativa de reforzadores obtenidos:

$$\frac{R_1}{(R_1 + R_2)} = \frac{r_1}{(r_1 + r_2)}$$

El resultado es muy robusto y se ha reportado en un sinnúmero de especies. A esta relación entre tasas relativas de respuesta y refuerzo se le conoce como la *ley de igualación* y en la última década del siglo pasado fue la ley más citada en la literatura psicológica. En la siguiente figura pueden verse los resultados de tres palomas: cada punto representa los datos de los últimos cinco días para cada par de valores de los programas de intervalo variable. La proporción de respuestas (tasas relativas) va de cero a uno. (Figura).

imagen

Igualación sería un resultado trivial, si por cada refuerzo hubiese solo una respuesta, sin embargo, el patrón de igualación en los refuerzos también se puede

obtener con un rango muy amplio de tasas relativas de respuesta que van más allá de la tasa específica del patrón de igualación.

Una forma más directa de estudiar la relación entre patrones de respuesta y patrones de refuerzo en protocolos de elección recurrente, es estudiando el **tiempo** asignado por los organismos a las diferentes opciones de refuerzo, en lugar de estudiar el número de respuestas discretas asignadas a las distintas opciones. Desde este enfoque, Rachlin y Baum estudiaron el comportamiento de las palomas: para ello, emplearon un espacio rectangular con un piso conectado a interruptores que permite medir el tiempo que una paloma pasa en cada lado del espacio rectangular. En cada uno de los dos extremos del espacio experimental había un comedero que asignaba comida de acuerdo a programas independientes de refuerzo de intervalo variable. En este experimento también se encontró que el tiempo relativo asignado por el organismo a un lugar iguala el refuerzo relativo obtenido en dicho lugar.

$$\frac{T_1}{(T_1 + T_2)} = \frac{r_1}{(r_1 + r_2)}$$

11.1 Desviaciones de Igualación

La igualación de las tasas relativas de respuesta al valor de las tasas relativas de refuerzo es un fenómeno muy robusto cuando ambas opciones de respuesta son reforzadas de acuerdo a programas de intervalo variable, sin embargo, se han encontrado desviaciones respecto al patrón de igualación cuando uno de los programas de refuerzo se cambio a otra regla, o cuando se establecen distintos tipos de reforzadores para las dos respuestas. Baum () reconoció dos tipos de desviaciones de igualación: Introducir el ejemplo de "ver Salir a alguien con una cantidad de frijoles de diferente tipo". ¿Es el resultado de preferencias o de precios?

1. *Sesgos*. Si en una visita al supermercado les ofrecen probar, sin ningún costo, frijoles negros o bayos, algunos de Uds. preferirán la prueba de los frijoles negros. Dada esa preferencia, si compran frijoles y ambos tienen el mismo precio, comprarán los frijoles negros. Sin embargo, qué frijoles deciden comprar depende de la diferencia en su precio modelada por su preferencia. Cuando los reforzadores para las dos respuestas son diferentes, por ejemplo, cuando una de las dos variedades de frijoles negros o bayos les brinda mayor satisfacción debido a su sabor particular, es posible que exista una preferencia por uno de ellos: esta preferencia tendrá un impacto sobre cada combinación de razones de refuerzo. Cuando esta razón es igual pero todavía se presentan diferencias en la tasa relativa de refuerzo, esta diferencia es un indicador del sesgo del organismo en favor de uno de los reforzadores. Los resultados se verían como los de la figura x: en ella, la

tasa relativa de respuesta se aleja de 0.5, aún cuando la tasa relativa de refuerzo es igual para las dos opciones.

imagen

2. *Sensibilidad.* En una misma visita al supermercado, un mismo producto que desean comprar es ofrecido por dos marcas distintas a precios diferentes: uno cuesta $11.00 y el otro $5.50. La diferencia en precio es de 2 a 1, sin embargo, los valores numéricos son difíciles de discriminar y para algunos de Uds. esta diferencia será percibida como de 3 a 1, mientras que para otros, la diferencia se percibirá como de 1.5 a 1. Sensibilidad es una segunda desviación de igualación, que ocurre cuando los organismos no son linealmente sensibles a la diferencia entre las tasas de refuerzo. Esto puede deberse a distinciones en la importancia de las diferencias en el valor de las opciones o a la dificultad para discriminar entre ellas. La figura x muestra como se vería la relación entre tasas relativas de respuesta y de refuerzo bajo distintos valores de sensibilidad. Cuando la tasa relativa de respuesta no es muy sensible a las tasas de refuerzo para cada respuesta, observamos valores cercanos a la indiferencia (panel de la izquierda en la figura) y a este resultado se le conoce como sub igualación. Cuando la tasa relativa de respuestas sobrevalora las diferencias en las tasas de respuesta, observamos que se prefiere mayoritariamente la mejor opción (panel derecho en la figura) y a este fenómeno se le conoce como sobre-igualación.

En resumen, el sesgo hace referencia a la preferencia del organismo por una de las opciones, la cual tiene un efecto multiplicativo al de la tasa de ocurrencia de los reforzadores para determinar las tasas de respuesta. El otro factor de desviación respecto a igualación es la sensibilidad del agente ante las diferencias en las tasas de ocurrencia de los refuerzos de las distintas opciones. Vamos a asumir que el sesgo y la sensibilidad varían independientemente el uno del otro.

imagen

11.1.1 Ley generalizada de Igualación

Para modelar el sesgo y la sensibilidad, Baum propuso una extensión de la ley de igualación que se conoce como la *Ley Generalizada de Igualación* y que captura las dos clases de desviaciones revisadas en la sección anterior. Primero propuso expresar la ley en términos de razones entre las distintas respuestas y los distintos refuerzos; y no de proporción entre una respuesta y el total de las respuestas o entre un refuerzo y el total de los refuerzos:

$$\frac{R_1}{R_2} = \frac{r_1}{r_2}$$

Como un segundo paso, Baum propuso que la razón de refuerzo es transformada por el agente, como una función de potencia con dos parámetros, similar a la propuesta por S. S. Stevens (AÑO) en la psicofísica sensorial.

$$\frac{R_1}{R_2} = \alpha \left(\frac{r_1}{r_2}\right)^{\beta}$$

Donde el parámetro β representa que tan *sensible* es la razón de respuesta a los cambios en la razón de refuerzos y el parámetro α representa el *sesgo** en la preferencia por una alternativa sobre otra.

Una forma, visualmente más clara de ver la ecuación anterior, es su transformación logarítmica:

$$\log \frac{B_1}{B_2} = \beta \log \frac{r_1}{r_2} + \log \alpha$$

En la figura x podemos ver su comportamiento y en el simulador uds. pueden jugar con diferentes valores de los parámetros de sesgo y sensibilidad. Podemos ver que bajo la transformación logarítmica, la ecuación de potencia se convierte en una familia de líneas rectas en las que β es la pendiente de la función y α es su intercepto. Cuando la sensibilidad es igual a uno y no hay sesgo (es decir, $\alpha = 0$), la ecuación es la ley de igualación y la línea recta parte del origen. Los valores de α positivos o negativos generan líneas rectas paralelas a la recta de igualación. Los valores del parámetro de sensibilidad β menores a uno representan sub-igualación y los valores mayores a uno representan sobre-igualación. En el primer caso, la importancia de la diferencia entre los refuerzos se empequeñece psicológicamente y en el segundo caso la misma diferencia se agranda.

imagen

En el marco de referencia de los modelos de elección basados en el valor de las consecuencias, la ecuación generalizada de igualación resulta ser una instancia de una regla de respuesta en la que la probabilidad de cada respuesta es una función de la diferencia entre reforzadores:

$$P(a_1) = F(\lambda(Qa_1 - Qa_2))$$

En nuestro caso, el parámetro β es λ y la función F es una función logística aplicada a la diferencia de los logaritmos de los dos refuerzos:

$$\log \frac{R_1}{R_2} = \beta(\log r_1 - \log r_2) + \log \alpha$$

La ecuación generalizada de igualación puede emplearse para distintos usos: ya sea evaluar la preferencia entre diferentes refuerzos (representada por el valor

del parámetro alfa); investigar bajo qué condiciones se obtiene igualación perfecta, es decir, cuando $\alpha = 1$ y $\beta = 1$ o finalmente, establecer la forma en la que las diferencias en refuerzo son transformadas en distintas distribuciones del comportamiento. El parámetro β puede interpretarse en al menos dos formas, primero como una propiedad del sistema: de la misma forma que el exponente para las funciones psicofísicas varía dependiendo de la dimensión sensorial, la sensibilidad beta podría variar dependiendo del tipo de refuerzo. En segundo lugar, beta también puede interpretarse como el resultado de un conjunto de manipulaciones experimentales y restricciones perceptuales del sistema que imponen límites en la discriminabilidad de las diferencias de refuerzos. Para dar un ejemplo de estas dos interpretaciones plausibles ante un único fenómeno de discriminación: consideremos la diferencia en el refuerzo que generan dos sabores placenteros distintos, esta diferencia es mucho más fácil de distinguir que aquella los refuerzos que otorgan dos sonidos melódicos distintos. Esta diferencia en la facilidad o dificultad para discriminar entre dos estímulos podría deberse a una propiedad intrínseca de los sabores como fenómeno respecto a los sonidos (interpretación 1), aunque también podría deberse a que nuestro sistema sensorial (el sistema sensorial humano) tiene una mayor facilidad para distinguir refuerzos por vía de la modalidad gustativa que por vía de la modalidad auditiva (interpretación 2). La segunda interpretación sugiere que esta dificultad discriminativa podría no presentarse en otras especies con sistemas sensoriales distintos, por ejemplo, en los murciélagos, que poseen sistemas auditivos más agudos que nosotros.

11.1.2 Igualación como un Mecanismo Adaptable

Una respuesta común al principio de igualación es considerarlo como una instancia de un mecanismo de adaptación, seleccionado para maximizar el total de refuerzos disponibles. Para atender esta posibilidad, es necesario conocer la función que relaciona, por un lado, a la suma del total de los refuerzos, y por el otro, a las distribuciones relativas de respuesta. Con ello, se puede evaluar si el patrón de respuesta de igualación corresponde al máximo de la función.

$$r_{total} = f\left(\frac{R_1}{R_1 + R_2}\right)$$

Dado que en programas de IV los refuerzos no se cancelan hasta que se obtienen, al animal le conviene seguir visitando ambas opciones y de esa forma obtener todos los reforzadores posibles. Sin embargo, el número de posibles distribuciones de respuestas que garantizarían obtener todos los refuerzos es enorme. El organismo podría hacerlo simplemente alternando constantemente cada respuesta (0.5) o pasando casi todo el tiempo en una de las opciones con una ocasional visita a la opción no atendida. Lo sorprendente es que dentro de ese amplio rango de posibles tasas relativas de respuesta que producen maximización, lo que se observa empíricamente es la proporción de respuesta a cada

opción que iguala la tasa de refuerzo relativa. La importancia de la igualación es que representa la solución observada, en equilibrio, a la multiplicidad de formas de maximizar el refuerzo total en programas concurrentes de intervalo variable.

11.1.3 ¿Es Maximización el Mecanismo que Subyace a Igualación?

Una pregunta muy diferente a la de si la igualación es un comportamiento adaptable es la de si, bajo condiciones de equilibrio, la maximización de la tasa de reforzamiento global es el "mecanismo" que guía el comportamiento del organismo y el cual subyace al patrón de respuestas observado en igualación. En otras notas veremos modelos en los que se maximizan diferentes variables, pero en estas nos concentramos en la maximización del número de refuerzos totales. Para comprender la pregunta, es necesario considerar que los fenómenos de maximización e igualación implican que los algoritmos que los organismos computan son diferentes para cada modelo. De acuerdo a esta versión de la maximización como mecanismo subyacente a la igualación, el algoritmo no distingue entre las dos respuestas disponibles y el refuerzo asociado con cada una de ellas: en lugar de ello, este solo computa y actualiza dos variables, la suma de refuerzos y la tasa relativa de respuestas. Noten que, debido a esto último, este modelo de acción no es una instancia de un modelo de elección basado en el valor de las respuestas individuales. En cambio, este modelo asume que los organismos cuentan con solo dos contadores, uno para la tasa relativa de respuestas y otro para la suma de los reforzadores obtenidos por las dos respuestas, sin distinguir entre su origen. Un reloj acumula el tiempo total T, durante el cual las dos respuestas se encuentran disponibles. El resultado del contador del total de refuerzos se divide entre el tiempo T. En estos casos, el organismo busca acceder al mayor número de refuerzos por unidad de tiempo. Este número de {r/T} representa la *ganancia* asociada con cada distribución posible de respuestas. Además, dentro de un proceso de ascenso de colina, como el visto en las notas x, la tasa relativa de respuesta se mueve en la dirección de una mayor tasa global de refuerzo hasta alcanzar un máximo, el cual puede ser local.

11.1.4 Igualación y Rentabilidad de las Respuestas

Otra explicación que puede dar cuenta del patrón de respuesta de igualación es que los organismos buscan igualar la rentabilidad de sus respuestas o tiempos. La rentabilidad es el número de refuerzos que se obtienen por tiempo o respuestas invertidos en una opción. Cuando ustedes deciden entre planes de ahorro bancario, la primera pregunta que hacen es cuál es la tasa de interés anual, lo que les permite saber cuánto ganarán anualmente por cada $1,000 pesos depositados en su cuenta. Igualación sugiere que esto es exactamente lo que hacen los agentes con la asignación de sus respuestas y tiempos: específicamente, igualación plantea que la regla que siguen los agentes es distribuir sus

respuestas y tiempos de tal forma que, en equilibrio, las dos opciones tengan la misma rentabilidad. En concreto, bajo este modelo computacional, el organismo en esencia registra la tasa de refuerzo asociada con cada opción de respuesta disponible y luego distribuye sus respuestas proporcionalmente en cada opción para igualar las rentabilidades. Igualación para respuestas y tiempos también puede expresarse en forma de razones, esto es, en lugar de hablar de una frecuencia relativa de 6 de 8 (0.75) refuerzos para una respuesta, hablamos de una razón de 6 a 2 (3) refuerzos para esa respuesta:

$$\frac{T_1}{T_2} = \frac{r_1}{r_2}$$

$$\frac{R_1}{R_2} = \frac{r_1}{r_2}$$

Reacomodando términos:

$$\frac{r_1}{T_1} = \frac{r_2}{T_2} \text{ y } \frac{r_1}{R_1} = \frac{r_2}{R_2}$$

Esta nueva forma de expresar la ley de igualación refleja que lo que se iguala es la rentabilidad de las distintas opciones de respuesta y/o tiempo. De este modo, lo que los organismos igualan son lo que podemos llamar como *tasas locales de refuerzo*. Así, cuando un organismo sigue la ley de igualación, este experimentará tasas iguales de reforzamiento local en todas las opciones disponibles. Es decir, si el organismo recibe un reforzador por cada 30 respuestas a la Opción A, este ajustará su número de respuestas a la Opción B para igualar la tasa de reforzamiento local de un reforzador por cada 30 respuestas. De manera similar, si el organismo recibe un reforzador por cada 30 segundos dedicados a la Opción A, este ajustará sus respuestas para recibir un reforzador por cada 30 segundos dedicados a la Opción B.

Es importante recalcar que para igualar las tasas de refuerzo local (la tasa de refuerzo por respuesta o por unidad de tiempo) el número de respuestas emitido por el organismo a las distintas opciones puede variar sustancialmente.

Por ejemplo:

-Si la Opción A brinda 1 reforzador por cada 30 respuestas y la Opción B brinda 1 reforzador por cada 60 respuestas, igualar las tasas de refuerzo locales requeriría que el organismo respondiera dos veces más a la Opción B que a la Opción A.

-Si distintos programas IV concurrentes se encuentran operando (por ejemplo, IV 30 s para la Opción A e IV 60 s para la Opción B), igualar las tasas de refuerzo local implicaría asignar distintas cantidades de tiempo y respuestas a cada opción.

El organismo esencialmente ajusta su comportamiento para obtener un "mismo rendimiento" sobre el tiempo/energía que invierte en todas las alternativas, lo que frecuentemente resulta en una distribución bastante desigual de respuestas.

También, vale la pena notar que igualdad en las tasas de reforzamiento locales (rentabilidad) ocurre naturalmente a través de la distribución del comportamiento del organismo, independientemente de los diferentes programas de intervalo variable programados para cada opción.

Computacionalmente, considerar que el algoritmo de igualación opera dentro de programas concurrentes de dos respuestas implica que los agentes tienen cuatro contadores, dos para respuestas y dos para reforzadores, y adicionalmente, disponen de dos relojes que se echan a andar cuando cambian a una de las opciones y que se detienen cuando regresan a la opción visitada anteriormente. Estos relojes se usan para computar las tasas locales de refuerzo.

Aqui falta el ejemplo numerico.

11.1.5 Maximización vs Rentabilidad

¿Es posible distinguir entre estas dos interpretaciones de la elección en programas concurrentes? Consideren el siguiente escenario. Una estudiante es la única heredera de dos tías de edad avanzada. El monto de la herencia que le deja una de ellas depende del número de visitas que la sobrina le haga; por otra parte, la cantidad que le deja la otra tía tiene un tope máximo y solo depende de que ella la visite ocasionalmente. El escenario de las tías ilustra un programa concurrente, con una de las opciones reforzada con un programa de intervalo variable y la otra con un programa de razón variable. En este escenario, la estrategia óptima de la sobrina es asignar la mayor parte de sus visitas a la tía más demandante, la que ejemplifica un programa de razón, y visitar ocasionalmente a la tía que ejemplifica el programa de intervalo.

Herrnstein y Heyman (), llevaron al laboratorio el escenario recién descrito. Expusieron a las palomas a programas concurrentes razón variable - intervalo variable. En este arreglo, una de las respuestas es reforzada de acuerdo a una regla temporal (programa de intervalo variable), mientras que el refuerzo para la otra respuesta depende del número de ellas (programa de razón variable).

Para entender la lógica del experimento, se debe tener presente que en programas de razón variable, la rentabilidad de la respuesta es el número de respuestas necesario para obtener un refuerzo. Por ejemplo, la rentabilidad de un programa de razón variable 30 para una de las opciones es un reforzador por cada treinta respuestas invertidas (1/30). De acuerdo a la ley de igualación, en un programa concurrente RV30 - IVx, el número de respuestas por refuerzo dentro del programa de intervalo debe ser también de 30. Sin embargo, de acuerdo a una regla de maximización global, la distribución óptima sería responder mayoritariamente en la opción reforzada con el programa de razón y visitar ocasionalmente la opción reforzada de acuerdo al programa de intervalo.

La siguiente figura muestra los resultados obtenidos por Heyman y Herrnstein. Puede verse que la distribución de respuestas de las palomas iguala la distribución de refuerzos, con un sesgo en favor del programa de razón. El patrón de igualación obtenido en este programa favorece al algoritmo de rentabilidad por encima del de maximización. Experimentos más recientes confirman este resultado y resaltan la utilidad de separar el sesgo por una opción (en este caso la opción RV) de la sensibilidad de los organismos hacia las diferencias en refuerzo de los dos programas.

Chapter 12

Comportamiento de Elección: Maximización Local

En el capítulo anterior, vimos que estudiando una medida agregada de respuestas y refuerzos, en equilibrio, la tasa relativa de respuestas tiende a igualar a la tasa de refuerzo que produce cada opción. Contemplamos dos posibles explicaciones computacionales para este fenómeno: la igualación de la probabilidad de refuerzo para las opciones de respuesta (igualación de la rentabilidad de las respuestas) y la maximización de la tasa global de refuerzo.

Una alternativa a estos modelos molares son los modelos de *maximización local*, los que asumen que igualación es el resultado de que en cada oportunidad de respuesta, los organismos eligen aquella respuesta asociada con el valor más alto de alguna variable local. Vamos a revisar tres miembros de esta familia de modelos, que se distinguen entre ellos por la variable local que cada uno maximiza:

- *Maximización Momentánea*: Modelos de elección basados en las probabilidades instantáneas de refuerzo de cada alternativa.
- *Mejoramiento*: Modelos de elección basados en las tasas locales de refuerzo asociadas con cada alternativa.
- *Valor Q de la respuesta.*

12.1 Maximización Momentánea

Este modelo asume que los organismos computan las probabilidades locales de refuerzo asociadas con cada respuesta. Consideren un programa concurrente IV

- IV. En esos programas, la probabilidad de un refuerzo para una respuesta en cada tecla es una función del tiempo transcurrido desde la última visita a una de ellas. Conforme incrementa el tiempo de estancia en una opción, incrementa la probabilidad de un refuerzo para la respuesta en la opción alterna. Este modelo fue propuesto por Shimp en 1969, quien argumentó que el cambio de una opción a otra era controlado por los cambios en la probabilidad de refuerzo asociados con el tiempo transcurrido en cada opción. Para él, la igualación global era el resultado de una regla de respuesta seguida por el organismo en la cual este respondía a la tecla que tuviera la mayor probabilidad de refuerzo al momento de la elección. Simulando este modelo, Shimp encontró que a nivel global este algoritmo resulta efectivamente en tasas de respuesta iguales a las tasas de refuerzo: reproduciendo así el patrón de igualación.

12.1.1 Evaluación experimental del modelo de maximización momentánea

12.1.1.1 1. Programas de refuerzo concurrente IV - IV de ensayos discretos

Empíricamente, el modelo de maximización instantánea predice regularidades en la estructura de los cambios del organismo de una opción a otra después de ciertas secuencias de respuesta. Si se computan las probabilidades de refuerzo de cada una de las opciones de respuesta, es posible determinar cuál respuesta debe seguir después de una secuencia de n respuestas consecutivas en una misma opción. La forma más accesible de estudiar esta predicción es empleando procedimientos de programas de refuerzo concurrentes IV - IV, pero de ensayos discretos. En estos programas, al animal se le presenta la oportunidad de elegir entre dos alternativas con una sola respuesta. Las dos opciones se presentan durante un breve periodo de tiempo. Después de una respuesta a alguna de las dos alternativas, inicia un breve intervalo entre ensayos sin opción de respuesta: al final de este intervalo, al organismo se le presenta una vez más una breve oportunidad para elegir entre una de las dos respuestas. Lo que es importante tener presente frente a estos protocolos es que los programas de refuerzo siguen corriendo durante los intervalos entre ensayos como si no hubiese discontinuidades en el tiempo. Al igual que en los programas IV - IV que hemos revisado, una vez que un refuerzo está disponible para una de las opciones, el reloj del programa IV se detiene y el refuerzo se guarda hasta que es recogido por el animal. Con este protocolo, mientras mayor es el número de elecciones repetidas por una de las opciones, mayor es la probabilidad de que un refuerzo esté esperando en la opción alterna.

En experimentos con programas de ensayos discretos, se mide la probabilidad de cambiar de alternativa después de varias secuencias de respuesta a una misma tecla. Mientras mayor sea el número de respuestas seguidas a una misma tecla, mayor será la probabilidad de refuerzo asociada con la respuesta alterna. De

acuerdo al modelo de maximización instantánea, debería observarse que la probabilidad de cambiar de alternativa es una función del número de respuestas que se hayan dado a la misma tecla. Por ejemplo, si el programa es uno concurrente de tecla verde con IV 1' - tecla roja con IV 3', el modelo predice la siguiente secuencia de tres respuestas: Verde-Verde-Roja. Por los valores de los programas de intervalo, después de cada dos respuestas seguidas a la opción verde resulta más probable que se otorgue un refuerzo a la opción roja.

En el experimento más citado con este procedimiento, Nevin () no encontró evidencia en favor del modelo de maximización instantánea. Ver Figuras.

imagen

Obsérvese en la figura del panel izquierdo que la probabilidad de refuerzo de la Opción Roja aumenta con la acumulación de elecciones de la Opción Verde por parte del organismo. Al mismo tiempo, en la figura del panel derecho, nótese que la probabilidad de cambiar a la Opción Roja no aumenta con la acumulación de elecciones de la Opción Verde por parte del organismo. Nota: la probabilidad de cambiar de opción se calculó con base al número de oportunidades para cambiar de opción en cada secuencia (run length). Así, estos resultados rompen con el patron de elecciones esperado según el modelo de maximización instantánea.

En un análisis posterior de su experimento original y de los datos de un experimento de Silberberg (), Nevin () encontró que la perseverancia en las opciones de respuesta (independientemente de su probabilidad de refuerzo) era el patrón más frecuentemente observado en ambos experimentos. Los animales tendían a quedarse en la tecla a la que habían respondido anteriormente y no a cambiar como una función del número de respuestas a esa tecla Ver fig.

imagen

12.1.1.2 2. Programas de refuerzo concurrente RV - IV de ensayos discretos

Un protocolo adicional para evaluar el modelo de maximización local es un programa concurrente de ensayos discretos RV - IV. En estos programas, la probabilidad de refuerzo para las respuestas asociadas a la opción RV es constante, mientras que la probabilidad de refuerzo para la respuesta asociada al programa IV cambia como una función de la última respuesta a esa opción. Una estrategia consistente con el modelo de maximización instantánea consiste en responder a la opción asociada con el programa RV inmediatamente después de recibir un refuerzo en la opción asociada con el programa IV. Este es el momento con menor probabilidad de refuerzo para la opción IV. De igual forma, la probabilidad de un cambio de la opción RV hacia la opción IV debe incrementar como una función del número de respuestas que se han dado a la opción RV.

La siguiente figura x presenta los resultados obtenidos por Williams () usando el protocolo anterior. En primer lugar, Williams encontró que los animales igualaban la frecuencia relativa de respuestas a la frecuencia relativa de refuerzos,

pero de manera aún más importante para estas notas: no encontró evidencia de que la respuesta fuese controlada por la probabilidad instantánea de refuerzo. El panel izquierdo de la figura muestra la probabilidad de refuerzo como una función del número de ensayos desde la última respuesta a la opción IV. Puede verse que la probabilidad de refuerzo para un cambio a la tecla IV es creciente; consecuentemente, la probabilidad de una respuesta a la tecla IV también debería incrementar como una función del número de ensayos desde la última respuesta a la opción IV. Sin embargo, el panel derecho de la figura muestra que la probabilidad real u observada de una respuesta al IV es constante o decreciente.

imagen

12.1.2 Conclusiones acerca del modelo de maximización Instantánea

De los experimentos con el protocolo de ensayos discretos recién descritos y muchos otros que no presentamos se pueden alcanzar las siguientes conclusiones:

- Consistente con la ley del efecto, la probabilidad de que se repita una opción de respuesta incrementa si esta es seguida por un refuerzo.
- También se observa un *efecto de perseverancia*: es más probable que se repita la respuesta a una opción que ya fue elegida en el pasado, independientemente de si esta respuesta es reforzada o no.
- Es importante agregar que la consistencia de los resultados descritos también depende de otras variables como el intervalo entre opciones de elección, el cual afecta la memoria del organismo sobre sus elecciones previas.

12.2 Modelo de Mejoramiento

El modelo de mejoramiento de Herrnstein y Vaughan () describe la dinámica de elección en situaciones de elección recurrentes, momento a momento. Como otros modelos de elección basados en el valor de las opciones, el modelo de mejoramiento cuenta con dos componentes:

- El primer componente es la especificación de la variable de decisión a partir de la cual el organismo elige (es decir, la variable que guía la elección de los organismos; ejemplos de estas variables son la probabilidad de refuerzo, el tiempo transcurrido en una opción o las tasas locales de refuerzo, entre otras opciones).

- El segundo componente es *la regla de respuesta*: esto es, con qué criterio elige el organismo.

Para el modelo de mejoramiento, la variable de decisión son las tasas locales de refuerzo, a las cuales previamente llamamos como la rentabilidad de las opciones. Esta variable se computa dividiendo el número de refuerzos para cada opción entre el tiempo asignado a cada una de ellas:

$$\frac{r_i}{T_i}$$

La rentabilidad puede computarse también para respuestas:

$$\frac{r_i}{R_i}$$

Asumiendo un tiempo total fijo, de acuerdo al modelo de mejoramiento, cada incremento en una unidad de tiempo asignada a una opción tiene como consecuencia la actualización de las dos tasas de refuerzo locales. Al incrementar el tiempo asignado t_i a una de las opciones, la tasa local de refuerzo de esa opción disminuye (dado que el denominador de la rentabilidad de esa opción crece); simultáneamente, este cambio también reduce el tiempo t_2 asignado a la otra opción (dado que el tiempo total es fijo), lo cual incrementa la tasa de refuerzo local asociada con la segunda opción (puesto que el denominador de esta segunda rentabilidad se achica).

El segundo componente del modelo de mejoramiento, la regla de elección, es una variante de maximización que consiste en seleccionar la opción con la mejor tasa de refuerzo local dentro de cada oportunidad.

Este modelo describe la elección dentro de un programa concurrente que se construye como un sistema dinámico y de retroalimentación. Es decir, bajo este arreglo, la opción de respuesta que tiene la mayor tasa local de refuerzo va cambiando como una función de la asignación de tiempos y respuestas por parte del organismo. Es por ello que este modelo puede dar cuenta de la dinámica global de acercamiento del organismo al patrón de igualación observado en programas concurrentes IV - IV. Para entender al modelo, tengan presente que al elegir la opción con la mejor tasa de refuerzo local, el organismo está incrementando el tiempo de estancia en esa opción: lo cual aumenta la base temporal t para computar la tasa de refuerzo local de esa opción. Simultáneamente, y dado el tiempo total fijo, este comportamiento reduce el tiempo de estancia en la segunda opción, lo cual incrementa la tasa local de refuerzo de esa segunda opción. De esta forma, cada aumento de una unidad de tiempo de estancia en la mejor opción por parte del organismo, tiene dos consecuencias: reduce la tasa de refuerzo local de esa opción e incrementa la tasa de refuerzo local de la otra opción. La transición del organismo de una opción a la otra ocurre cuando la dirección de la diferencia entre las dos tasas cambia de signo y, básicamente, la opción alterna se vuelve mejor que la opción actual. Pueden observar que en estos programas, el modelo de mejoramiento es uno de retroalimentación y corresponde a un algoritmo cuyo blanco es la reducción de la diferencia entre

las dos tasas de refuerzo locales; el seguimiento de este algoritmo a la larga (o en equilibrio) resulta en la igualación global de las tasas de respuesta a las tasas de refuerzo para ambas opciones (el patrón de igualación). Consideren como ejemplo un programa concurrente IV 1 min - IV 2 min. Supongan que al inicio de la sesión de una hora, el organismo asigna la mitad de su tiempo a cada alternativa. Supuesto importante: Vamos a asumir que el animal responde a una tasa moderada que garantiza que en ambos programas de IV el organismo obtendrá el máximo número de reforzadores posibles (para 1 hora), independientemente del tiempo que efectivamente dedique a cada opción durante la sesión. Esto ocurre porque en los programas de intervalo variable, los reforzadores se "acumulan" durante el tiempo que el animal no está respondiendo a esa opción. Por lo tanto, aunque el animal dedique menos de 1 hora a una opción, si este responde de manera suficientemente rápida, efectivamente podría obtener todos los reforzadores correspondientes a 1 hora pero en una menor cantidad de tiempo. Dado este supuesto, las tasas de refuerzo locales para las dos opciones se calculan como el número máximo posible de refuerzos en una hora dividido por la proporción de tiempo realmente asignada a cada opción. Así, la tasa local de refuerzo para la tecla IV 1' sería de 60 reforzadores máximos / 0.5 hr = 120 refuerzos por hora. Una vez más, esto significa que aunque el animal solo dedica media hora a esta opción, este puede obtener los 60 reforzadores disponibles durante la sesión completa gracias a la acumulación propia de los programas IV. Nótese que si el organismo le asigna toda la hora a responder a este programa, su tasa local de refuerzo será de 60/1 hr = 60 refuerzos por hora. Al reducir el tiempo asignado a esa opción a media hora (0.5 hr), la rentabilidad de esa opción aumenta a 120 refuerzos por hora, aunque el número absoluto de reforzadores (60) permanece constante. Para el IV 2', la tasa local de refuerzo es de 30 reforzadores máximos / 0.5 hr = 60 refuerzos por hora. Dados estos resultados y de acuerdo al modelo de mejoramiento, el animal debería escoger asignar más tiempo a la opción con el programa IV 1' que a la opción con el programa IV 2', ya que la primera presenta una mayor tasa de refuerzo local (120 vs 60). Supongan ahora que como resultado de asignar más tiempo a la opción IV 1', el organismo termina dedicando el 90% de su tiempo a esta opción. Ahora, para la opción asociada con el IV 1 min, la tasa de refuerzo local sería de 60/0.9 hr = 66.7 refuerzos por hora. En el IV 2 min, la tasa de refuerzo local sería de 30/0.1 hr = 300 refuerzos por hora. Recordamos que los programas en sí no han cambiado: lo único que está modificando la rentabilidad percibida de ambas opciones es la manera en la que el organismo distribuye su tiempo entre ellas. Pueden observar que al incrementar el tiempo asignado al IV 1', paradójicamente se incrementó la tasa local de refuerzo asociada con la opción IV 2' (de 60 a 300 refuerzos por hora) y por lo tanto, si el organismo sigue la regla de mejoramiento, su subsecuente elección debe ser la opción del IV 2'. Bajo este modelo, el equilibrio se alcanza cuando la proporción de tiempo asignado a las dos opciones por parte del organismo es de 2/3 para IV 1' (60/0.666 90) y 1/3 para IV 2' (30/0.333 90). Lo anterior es aproximadamente igual a "90 refuerzos por hora" en ambas alternativas. Así, en el punto de equilibrio, ambas alternativas ofrecen exactamente la misma tasa de refuerzo local (90 por

hora), lo que explica por qué el organismo deja de cambiar su distribución de tiempo entre las opciones tras alcanzar este punto. Este resultado es equivalente a la **igualación** global de tasas relativas de respuesta a las tasas relativas de refuerzo: 0.66 hr/(0.66 hr + 0.33 hr) 60 r/(60 r + 30 r) En el correspondiente simulador, ustedes podrán ver la dinámica del sistema para diferentes valores de programas concurrentes.

Un problema importante del modelo de mejoramiento es que deja sin especificar la ventana temporal que los organismos requieren para computar las tasas de refuerzo locales de las distintas opciones. ¿Estas tasas se estiman hasta finalizar la duración de cada sesión experimental? O bien aún, ¿seguirá el organismo algoritmos menos evidentes? Por ejemplo, ¿reiniciar la computación de las tasas para las dos opciones cada 10 minutos?¿O borrar la historia con las opciones de respuesta experimentadas días atrás? Las respuestas a todas estas preguntas no son evidentes. Y al mismo tiempo, estas tienen importantes implicaciones para la aplicación del modelo a entornos volátiles, los cuales frecuentemente cambian las condiciones de refuerzo para las diferentes opciones a lo largo del tiempo. Este tema será abordado en otras notas.

12.3 Modelo de Valor Q de la respuesta

De los modelos que dan cuenta de la elección recurrente, el modelo de valor Q que vimos en las notas x es el más cercano a la ley del efecto original planteada por Thorndike. La variable de decisión de este modelo es el valor Q adquirido por cada opción de respuesta. Este valor representa la integración de la historia de reforzamiento de cada opción que resulta de la regla del error de predicción. Por otra parte, la regla de respuesta bajo este modelo consiste en la elección probabilística de la respuesta con mayor valor Q en cada oportunidad. Presten atención a que en este modelo, cada respuesta adquiere su valor exclusivamente como función de los refuerzos que produce y es independiente del valor de las demás alternativas presentes.

Por consecuente, una predicción importante de este modelo es que cuando al organismo se le presenta la oportunidad de elegir entre dos respuestas con diferentes valores Q, su elección debe ser independiente del contexto donde fue adquirido el valor Q de cada opción. Imaginen que en su ciudad hay tres cadenas de cafeterías (A, B y C). Cerca de su casa y cerca de su trabajo, hay dos sucursales disponibles (digamos que A y B están cerca de su casa; y B y C están cerca de su trabajo). Una de las cadenas de cafetería, la B, tiene una sucursal en los dos escenarios donde ustedes compran café (a la sucursal cerca de su casa le llamaremos B' y a la sucursal cerca de su trabajo le llamaremos B' '). Las sucursales de la cadena B tienen un logo que las distingue. Digamos que sus menús son similares, por lo que las dos sucursales B, aunque tienen sutiles diferencias en cuanto a su staff, comparten a grandes rasgos el mismo programa de recompensas (B' = B' '). En cambio, las cafeterías A y C son claramente dis-

tinguibles en cuanto a calidad con relación a ambas cafeterías B. En resumen: A es mucho mejor que B (A > B) y C es mucho peor que B (C < B). Sin embargo, un domingo ustedes acuden a otra zona de la ciudad y se dan cuenta de que las dos sucursales de las cafeterías B (B' y B'') han sido reubicadas ahora en una misma y nueva zona de la ciudad. Dado que los programa de recompensas de ambas opciones son iguales (B' = B''), ustedes deberían ser indiferentes entre ellas. Sin embargo, es posible que los valores percibidos de B' y de B'' dependan de cuál era el otro restaurante con el que ambas opciones competían dentro de su contexto previo (A o C).
Una forma de evaluar experimentalmente esta predicción sobre la relevancia del contexto previo de los refuerzos es presentar a los animales con dos situaciones de elección diferentes: Programa concurrente 1: El animal puede elegir entre: Opción A: reforzada con intervalo variable IV(a) Opción B: reforzada con intervalo variable IV(b) Programa concurrente 2: El animal puede elegir entre: Opción A: reforzada con el mismo intervalo variable IV(a) que en la Situación 1 Opción B': reforzada con un intervalo variable IV(b') diferente al de la Situación 1 Posteriormente, se evalúa la preferencia del animal entre las opciones A de ambos programas concurrentes, las cuales comparten exactamente el mismo valor de reforzamiento (Q). Bajo este protocolo, cada par de opciones (cada componente o programa concurrente) se encuentra vigente en períodos de tiempo separados dentro de una misma sesión (formalmente a este arreglo se le conoce como programas múltiples de refuerzo). Una vez alcanzado el equilibrio en cada programa concurrente IV - IV: se presenta una nueva combinación de las opciones comunes a ambos programas. De acuerdo a los modelos de refuerzo, de mejoramiento y de valor Q, la elección en los periodos de prueba debe reflejar los valores adquiridos por cada opción individual.

Belke () corrió una versión de este protocolo. Para uno de los componentes de programas concurrentes (Componente A), los valores de las opciones eran Tecla Blanca (IV 20) vs Tecla Roja (IV 40), mientras que para el otro programa concurrente (el Componente B), las opciones eran Tecla Verde con IV 40 vs una Tecla Amarilla con IV 80. Así, para el Componente A, la opción del IV 40 ocurría en el contexto de una opción IV 20 (la Tecla Roja) que era dos veces más rica. Mientras tanto, para el segundo par de estímulos (el Componente B), la opción IV 40 (la Tecla Verde) se presentaba en el contexto de otra opción reforzada con un IV 80 (la Tecla Amarilla), la cual era dos veces menos rica. De acuerdo al modelo Q que no considera el contexto de los reforzadores, las dos opciones de respuesta reforzadas con el IV 40 (las Teclas Roja y Verde) en ambos programas deben tener el mismo valor. Sin embargo, otra posibilidad es que el valor adquirido por una opción de respuesta dependa del contexto de los refuerzos proporcionados a las respuestas alternativas presentes. En este caso, la Tecla Verde reforzada con el IV 40 dentro del contexto de otra opción reforzada con un IV pobre, adquiere un valor más grande que la Tecla Roja, también reforzada con un IV 40 pero que fue entrenada previamente en un contexto con una opción reforzada con un IV más rico.

Contrario a la predicción de indiferencia arrojada por el modelo de refuerzo

Q, Belke encontró que los animales sí preferían la opción IV 40 entrenada en el contexto pobre, por encima de la misma opción entrenada en el contexto rico. Esta evidencia sugiere que el impacto de los reforzadores sobre el valor de una respuesta depende de los reforzadores obtenidos por las otras opciones presentes en el pasado (es decir, con qué otros estímulos ha convivido cada estímulo previamente).

Los resultados anteriores son consistentes con la siguiente interpretación más local de la ejecución en programas concurrentes. La interpretación consiste en suponer que en estos programas lo que el animal aprende es el tiempo que debe pasar en una opción antes de cambiar a otra alternativa. Suponemos que estos tiempos se encuentran relacionados linealmente a la tasa de refuerzo en la tecla alternativa: así, mientras más pobre es el programa alterno, mayor será el tiempo que el organismo pasará en una alternativa. De esa forma, en la prueba de Belke, la preferencia por el "IV 40 que fue entrenado con el IV 80" respecto al "mismo IV 40 que fue entrenado con el IV 20" se explica porque en el primer programa concurrente (Componente B), el organismo aprendió que debía pasar más tiempo por visita en la opción IV 40, mientras que en el segundo programa (Componente A), este aprendió que debía pasar menos tiempo por visita en la opción IV 40.

12.4 Reflexiones Finales Sobre los Modelos de Elección Basados en Valor

- La igualación de tasas relativas de respuesta a tasas relativas de refuerzo es un fenómeno muy robusto.
- Igualación es el resultado de un proceso estabilizador de retroalimentación, gobernado por las propiedades temporales de los programas de intervalo.
- Igualación ilustra la importancia de entender a los programas de refuerzo como restricciones temporales o de respuesta sobre la distribución de comportamientos.
- En un nivel molecular, la dinámica del movimiento hacia igualación ilustra la relevancia del principio de refuerzo, entendido como un algoritmo de ascenso de colina, como el que vimos en las notas x. Este algoritmo postula que el sistema compara, en cada oportunidad de respuesta, el valor de las variables de decisión asociadas con cada alternativa y elige aquella con el valor más alto. La distribución del comportamiento entre diferentes opciones (en equilibrio) es el resultado de diversos factores: primero, el seguimiento del algoritmo de ascenso de colina por parte del organismo; segundo, los ajustes que sufre dicho algoritmo al operar bajo las diferentes restricciones impuestas por los distintos programas de refuerzo; y tercero, la regla de aprendizaje de los valores de la variable de decisión. *La importancia del tiempo de estancia reforzado en cada opción sobre la elección de los organismos.

- El avance en los modelos de elección requiere de la consideración de protocolos que capturen la incertidumbre y la volatilidad propia de los entornos naturales de los organismos, un tema que quedará para otra nota.

Referencias