

ON THE STRUCTURE AND ROLE OF OPTIMALITY MODELS IN THE STUDY OF BEHAVIOR

MARCO VASCONCELOS

INÊS FORTES

ALEX KACELNIK

2017

Dada la selección natural, en el largo plazo las características biológicas, incluyendo las conductuales, parecen haber sido diseñadas para maximizar la adecuación de los organismos. Asumir ese diseño óptimo permite generar hipótesis de los mecanismos de decisión.

¿No sería razonable anticipar que nuestro entendimiento de la mente sería ayudado de saber para qué fue diseñada? La palabra “diseño” se usa en un sentido no-teleológico y separado de la idea de *diseño inteligente*. El propósito del diseño es distinto de las metas propias del organismo particular. Por ejemplo, el apareamiento puede ser guiado individualmente por el deseo sexual, pero su propósito según la optimalidad biológica es maximizar la representación en las generaciones futuras.

La optimalidad entonces permite proponer cómo funcionan los animales y qué es importante para la selección. Pero se debe abandonar varios malentendidos: que los biólogos evolutivos esperan que los animales se comporten de forma perfecta, que los modelos de optimalidad requieren que los animales computen óptimos, y que la hipótesis a probar es si los animales son óptimos o no. En tanto el ambiente actual refleja al del pasado de la especie, su conducta será en promedio adaptativa.

COMPONENTS OF OPTIMALITY MODELS

Kacelnik y Cuthill argumentan que los modelos de optimalidad son un ensamble de al menos tres supuestos interconectados independientemente probables: *strategy set*, *feedback function*, y *currency*.

The strategy set

La optimalidad ve a toda conducta como elección (no necesariamente deliberación consciente) entre alternativas. Es decir, los modelos definen un rango de conductas potenciales o *strategy set* inspirado en la observación de los organismos (*e.g.*, los caballos eligen entre trotar u galopar; las aves, entre caminar y volar). Las limitaciones físicas son evidentes, pero las psicológicas no. Además, los modelos solo pueden producir como resultado un elemento de su propio *strategy set*. Esto no es una limitación, porque el *strategy set* es una hipótesis sujeta a prueba y mejora recursiva.

The feedback function

La función de realimentación es una hipótesis de qué le sucede al actor en función de lo que hace y de su estado actual (estado energético, riesgo de depredación, etc). El modelador hace conjeturas informadas para incluir un número razonable de relaciones acción-consecuencia que pueden influir en la psicología del organismo.

The currency

La adecuación Darwiniana es la contribución del organismo al *pool* genético de la especie. Sin embargo, esto no puede medirse en la escala conductual, de modo que los modeladores deben identificar variables medibles a corto plazo que tengan una relación clara con la adecuación a largo plazo, como la tasa de ingesta, vulnerabilidad a la depredación, o balance nutricional. También se trata de hipótesis a probar, de modo que si una moneda de cambio particular no parece ser maximizada por la conducta, nuevos modelos modificarán esta presunción.

PREDICTIONS OF OPTIMALITY MODELS

Los modelos generan predicciones y experimentos que las desafíen. Para predecir hay que preguntar qué miembro del *strategy set* maximiza la moneda de cambio dada la función de realimentación. Las predicciones pueden falsearse, lo que es informativo pues indica que al menos una de las hipótesis componentes está mal. La predicción nunca es que los animales son óptimos. La presunción de que la selección natural optimiza permite pasar a las ciencias conductuales de una aproximación descriptiva a una hipotético-deductiva.

Se intentará mostrar que la lógica subyacente en la ecología conductual es una herramienta poderosa e infravalorada por los psicólogos experimentales. La optimalidad complementa en lugar de competir con una aproximación mecanística.

PATCH EXPLOITATION: THE MARGINAL VALUE THEOREM

Al mantener una misma actividad por largo tiempo, las ganancias tienden a decrecer, de modo que es más beneficioso cambiar a otra actividad pasado cierto punto. Las estrategias de cambio entre actividades se han estudiado en el forrajeo, y uno de los modelos más estudiados es el teorema de valor marginal. Una situación en la que se aplica es en el forrajeo de lugar central (*central place foraging*), en donde un animal explora las cercanías de su nido para encontrar comida y regresar para entregarla a sus polluelos.

Dado que la tasa de entrega de comida afecta la adecuación de los polluelos es un buen candidato *a priori* para moneda de cambio. El *strategy set* serían todas las acciones bajo el control del ave, como la alocaión de comida entre sí mismo y sus polluelos, el tiempo para dejar de recolectar y volver al nido, y la distribución de comida entre polluelos. En el caso de cuándo dejar de recolectar comida, el *strategy set* son los posibles tiempos en el parche, y la función de realimentación es cuánto varía la moneda de cambio (tasa de aprovisionamiento) en función del tiempo en el parche. Una limitación es que la tasa de aprovisionamiento no incrementa linealmente con el tiempo dado que las presas acarreadas ralentizan al ave.

La moneda de cambio (tasa de aprovisionamiento $R(t)$) se puede expresar como

$$R(t) = \frac{G(t)}{\tau + t} \quad (1)$$

donde $G(t)$ es la curva de ganancias acumuladas en función del tiempo desde la llegada al parche, τ es el tiempo de viaje medio entre el nido y el parche, t es el tiempo entre

aterrizar en el parche y abandonarlo (tiempo de parche). Se debe buscar el valor de t que maximice $R(t)$ dada la forma de $G(t)$ y el valor de τ . Si la tasa de captura decrece con el tiempo en el parche (la segunda derivada de $G(t)$ es negativa), entonces el cálculo nos dice que el punto óptimo de t será cuando la primera derivada de R con respecto a t sea nula, suponiendo que en ese punto la segunda derivada sea negativa. Este valor t_{op} es el tiempo de parche predicho. Dado que el tiempo de parche y la carga se relacionan mediante $G(t)$, predecir t_{op} también especifica la carga óptima por viaje $G(t_{op})$.

Según este modelo la política que maximiza la tasa se obtiene de una regla matemáticamente simple: permanecer en el parche mientras la tasa local ($G(t)$), la primera derivada de $G(t)$, exceda la tasa general esperada R .

Un organismo solamente elige en dónde invertir su siguiente unidad de tiempo, pero este modelo incluye la capacidad de ajustar la conducta según la tasa global R en el ambiente mediante experiencia o mediante alguna clave ambiental. También incluye la capacidad de percibir la tasa local $G(t)$ mientras baja en función del tiempo de parche.

Mientras el tiempo de viaje τ incrementa, también lo hacen el tiempo de residencia óptimo y la carga óptima. Esta predicción se cumple en toda prueba experimental del teorema de valor marginal.

El teorema de valor marginal incluye varias simplificaciones:

1. $G(t)$ es tomado como función continua aunque las presas reales son discretas.
2. Los ciclos de forrajeo se asumen como idénticos, aunque en la realidad su variabilidad puede tener un efecto.
3. Hay una sola moneda R (tasa de aprovisionamiento), pero puede haber otras limitaciones como la depredación que impongan *trade-offs*.
4. El padre necesita también comer para permanecer vivo, lo que no se toma en cuenta.
5. No se toma en cuenta el costo metabólico del viaje.
6. No se menciona el mecanismo. El agente necesita adquirir y procesar la información relevante igual que el modelador, pero no se menciona cómo.

Todas las simplificaciones son susceptibles de ser refinadas mediante la experimentación, *e.g.*, Kacelnik aplicó el teorema de valor marginal a *starlings* alimentando a sus crías, y encontró que el modelo predecía datos sistemáticamente debajo de los reales. El modelo cometía una simplificación al tomar a todos los tiempos como equivalentes. Sin embargo, al considerar que el tiempo de vuelo es más costoso que el tiempo en un parche o en el nido, el ajuste del modelo incrementó significativamente.

El teorema de valor marginal no solo se utiliza en forrajeo, también puede aplicar a situaciones como el apareamiento (*e.g.*, las moscas de estiércol macho deben decidir cuánto tiempo permanecer con una hembra después de la cópula).

The role of psychology in optimal foraging

Para llenar el espacio de los mecanismos proximales dejado por el teorema de valor marginal, los ecólogos conductuales proponen reglas simples llamadas *reglas de dedo* capaces de producir conductas cercanas a los óptimos predichos. Se han propuesto múltiples reglas para situaciones específicas, como *hunting by expectation* (los animales abandonan los parches tras un número dado de capturas), *giving-up time* (se abandona el parche cuando pasa un tiempo dado desde la última captura), *patch-residence time* (se abandona un parche tras un tiempo de explotación dado), y hasta reglas de actualización Bayesiana. Estas reglas muestran que los animales pueden aproximar soluciones óptimas sin la necesidad de hacer los cálculos que hacen los investigadores. Pero estas reglas tienen limitaciones: (1) son específicas al dominio, por lo que no son aptas para animales en ambientes heterogéneos con múltiples tareas. Esto implicaría la existencia de una biblioteca extensa de reglas de dedo para cada situación, y un mecanismo de selección. (2) Esta aproximación ignora los mecanismos conocidos en el campo de la cognición y el aprendizaje animal, como el aprendizaje por reforzamiento. El algoritmo adaptativo podría ser un proceso de aprendizaje y no el uso de una regla específica para cada problema.

Se argumenta que la optimalidad es un marco de referencia que permite integrar hipótesis funcionales y mecanísticas. La función adaptativa limita qué mecanismos psicológicos evolucionan, del mismo modo que los mecanismos psicológicos de dominio amplio determinan la naturaleza de los problemas que cada animal resuelve. Como ejemplo se puede pensar en la incorporación de procesos psicológicos de *interval timing* en el forrajeo óptimo dada la sensibilidad al tiempo necesaria en los organismos. Dado que las propiedades de un mecanismo psicológico de timing no necesitan ser vistas como dedicadas a una única situación experimental, esta aproximación es preferible al uso de reglas de dedo.

Optimality and environmental variability

Hasta ahora se han mencionado modelos que lidian con promedios (en tiempos de viaje o ganancias), pero hay evidencia que indica que la variabilidad es determinante en la conducta: se ha encontrado que la explotación de parches es afectada por el tiempo de viaje más reciente, y que las palomas disminuyen el tiempo de parche cuando el tiempo de viaje es más variable.

La insensibilidad a la variabilidad predicha por el teorema de valor marginal indica que el tipo de forrajeador ideal que describe toma decisiones de acuerdo con las oportunidades promedio esperadas en el futuro, y no de acuerdo con los costos de viaje ya pagados. Un forrajeador ideal abandona un parche cuando espera, en promedio, una ganancia mayor en otro sitio. Si el ambiente tiene variabilidad aleatoria, entonces la tasa media esperable en el futuro es definida por la media. Sin embargo, el futuro solo puede anticiparse con el pasado, y un algoritmo sensible incluiría un mecanismo de ponderación en la recencia.

THE SELF-CONTROL PROBLEM IN INTERTEMPORAL CHOICE

El autocontrol en elección intertemporal también se beneficia de la integración de aproximaciones funcionales y mecanísticas. Se puede pensar en aves que deban llevar comida a sus nidos eligiendo entre dos parches que entregan comida de distinto tamaño tras distintos tiempos de búsqueda: pequeño inmediato y grande demorado. Dado que la moneda de cambio es la cantidad de comida entregada, las aves serían indiferentes cuando la razón entre tamaño y tiempo es idéntica, es decir, cuando la espera en la alternativa grande demorada es exactamente compensada por su magnitud, lo que es equivalente a

decir que hay descuento temporal.

Esto ha sido ampliamente estudiado por psicólogos experimentales. En sus estudios no se hace la pregunta típica de forrajeo óptimo: “¿qué relación entre tamaño y demora igualaría las tasas de ganancia?”, sino “¿qué función describe el valor de una recompensa como una función del tiempo de espera?”. Esta función suele llamarse *de descuento*. Con esta perspectiva *one shot* implícita no se toma en cuenta el costo de oportunidad, y se describen los resultados diciendo que los animales son impulsivos y rechazan ganancias a largo plazo por la inmediatez. Según este análisis, si una recompensa pudiese ser consumida en un tiempo cero, su valor se dispararía hacia el infinito.

En esta aproximación *one shot* hay explicaciones descriptivas y normativas del descuento. Las normativas sugieren que el descuento debería tener una función exponencial dado que la probabilidad de realmente obtener la comida declina con el tiempo, lo que se conoce como hipótesis de *discounting by interruptions*. Esta lógica contrasta con los tratamientos de forrajeo óptimo que ven al tiempo con un recurso que limita y se enfocan en múltiples ciclos de decisiones.

Una alternativa al descuento exponencial es el descuento hiperbólico de Mazur:

$$V_i = \frac{S_i}{1 + kt_i} \quad (2)$$

donde S_i es el valor subjetivo de la recompensa si estuviese inmediatamente disponible, k es un parámetro libre con dimensiones recíprocas al tiempo, y t_i es la demora entre el momento de evaluación y el resultado. Las diferencias individuales se explican mediante cambios en k .

Esta ecuación solamente le da importancia a la demora entre la elección y el resultado. Hay evidencia que indica que los intervalos entre ensayo tienen muy poco efecto en los experimentos de autocontrol, lo que es sorprendente dado que en problemas de explotación de parches el tiempo de viaje tiene un impacto fuerte y predecible. La respuesta a este dilema puede estar en la posición temporal de los componentes del ciclo con relación al momento en que el sujeto hace su elección. En explotación de parches, la decisión del forrajeador es cuándo abandonar el parche actual, por lo que el costo de viaje ocurre entre la decisión y sus consecuencias. En el paradigma de autocontrol la decisión es la elección entre pequeño inmediato y grande demorado, y la demora ocurre entre esa decisión y la consecuencia, con el intervalo entre ensayos colocado después de la consecuencia de la elección. Se propone que los animales son muy sensibles a los tiempos entre las decisiones y las consecuencias, pero relativamente insensibles a otros intervalos debido al problema de *asignación de crédito*: un animal atribuirá la responsabilidad de una consecuencia a su decisión anterior.

En una situación en la cual un forrajeador deba elegir entre dos alternativas que difieren en contenido energético neto S_i , tiempo de manejo t_i , tiempo de búsqueda τ_i , y tasa metabólica durante la búsqueda m_i , la tasa neta que resultaría de escoger solo la presa i sería:

$$NetR_i = \frac{S_i - m_i\tau_i}{\tau_i + t_i} \quad (3)$$

Esta ecuación se ha usado para predecir el comportamiento de *starlings* eligiendo entre

buscar comida caminando y volando y se obtuvo un buen ajuste con los datos, lo que indica que los animales sí incluyen los costos de tiempo y energía en sus preferencias por fuentes de comida.

En resumen, el descuento hiperbólico es la forma de descuento en elecciones intertemporales predicha por modelos de optimalidad. Se esperan desviaciones en situaciones en que las decisiones son influidas por reforzamiento. Las elecciones *one shot* no son en general modelos adecuados de la elecciones animales debido a que los animales aprenden los parámetros de las tareas mediante elecciones repetidas.

THE STRUCTURE OF FORAGING ENVIRONMENTS AND CHOICE

Se asume que las decisiones implican *tradeoffs*, por lo que es usual pensarlo como una comparación que implica esfuerzo cognitivo y tiempo.

Se propone partir de la idea de que los encuentros en la naturaleza son secuenciales y que las decisiones tomadas son entre perseguir y no hacerlo. Así, los animales hacen una valoración de cada fuente de recompensa cuando la encuentran mediante los mecanismos de aprendizaje por refuerzo. Esta valoración es una función de la redituabilidad recordada del tipo de presa con relación a la redituabilidad del ambiente entero. Cuando los organismos encuentran opciones singulares la valoración se expresa como latencia, que refleja la probabilidad de no perseguir la presa. Las latencias deberían bajar cuando aumenta la redituabilidad de la presa, y aumentar según suben las reservas energéticas del animal.

Dado el ruido aleatorio, los encuentros generan una distribución de latencias. Cuando dos alternativas se encuentran simultáneamente, ambas generan su propia distribución independiente y aquella con la menor latencia muestreada se manifiesta conductualmente. Así, no hay comparación deliberada entre las alternativas. Formalmente, la probabilidad P_A de elegir la opción A sobre B está dada por la probabilidad conjunta de que la latencia para A sea igual a x y la latencia para B exceda a x , integrada para toda posible x :

$$P_A = p(l_A < l_B) = \int_0^\infty f_A(x) \cdot [1 - F_B(x)] dx \quad (4)$$

donde l_A y l_B son muestras aleatorias de sus distribuciones, f_A es la función de densidad de probabilidad de las latencias para A , f_B es la función de distribución acumulativa de latencias para B , y x es un valor particular de latencia.

En resumen, la latencia para aceptar una opción es una función de tres variables: las propiedades de la opción, la tasa de ganancia promedio del ambiente, y el estado energético en el momento del aprendizaje (no en el momento de la elección).

Estas presunciones combinan la lógica de forrajeo óptimo con observaciones empíricas.

Una desviación con respecto a la optimalidad *a priori* es la existencia de latencias ante opciones singulares, pues se esperaría que una alternativa que será aceptada lo sea de inmediato. Aún así, se ha encontrado que esa latencia es dependiente de los parámetros mencionados.

Se ha argumentado que las latencias pueden ser un artefacto de la falta de elección en el laboratorio, pues las condiciones que las generan son las mismas que generan el rechazo de una alternativa en condiciones naturales. Aun así, es un hecho que las latencias existen y deben incluirse en el *strategy set* de los modelos. Al incluir las latencias se puede hacer la

predicción contraintuitiva de que las latencias en ensayos de elección serán menores que en ensayos singulares.

Esta predicción se opone a la ley Hick-Hyman, que indica que toda elección es un *tradeoff* entre precisión y costo temporal de evaluación.

Las elecciones simultáneas pueden predecirse de varias formas bajo la lógica del modelo de elección secuencial. Una de ellas, llamada *molar*, usa la distribución completa de latencias de encuentros secuenciales para predecir la proporción global de elecciones. La alternativa, llamada *molecular*, intenta predecir el resultado de cada elección con el encuentro secuencial más reciente de cada opción.

Es difícil probar la predicción del acortamiento de latencias debido a que tienen un límite inferior pero no uno superior, lo que resulta en efectos de piso. Además, ese acortamiento se muestra más en la opción menos preferida, por lo que los tamaños de muestra suelen ser pequeños. Aun así, se ha encontrado el efecto en varios experimentos.

No tiene sentido explicar la conducta únicamente recurriendo a explicaciones mecánicas o funcionales. Una actúa en la experiencia de los individuos, y otra en su pasado transgeneracional. Ambas son necesarias para una explicación completa. Incorporar mecanismos psicológicos conocidos en el *strategy set* de los organismos es una buena forma de enlazar las dos aproximaciones.

CONCLUSION

Dado que la conducta y los procesos cognitivos son productos de la evolución, la investigación conductual se puede beneficiar del modelamiento de optimalidad que, aunque sigue una lógica evolutiva, está apoyado por hipótesis psicológicas específicas que pueden probarse experimentalmente.

Los modelos en sí mismos están formado de hipótesis falseables sobre *strategy set*, *feedback function* y *currency*.

Los ambientes de prueba rara vez se ajustan a los ambientes en que evolucionaron los mecanismos conductuales. Esto se puede abordar en experimentos de selección artificial. Por ejemplo, se ha logrado predecir los parámetros de aprendizaje a lo largo de solo 30 generaciones de moscas de fruta. Pero aun cuando las predicciones pueden fallar debido al desajuste de ambientes, los modelos dan un marco fuerte para el estudio de mecanismos conductuales al proporcionar posibles equivalentes naturales para los protocolos experimentales.

Una visión integrativa de la conducta y la cognición debe combinar aproximaciones evolutivas y psicológicas.