

EXPLORATORY SEARCH: INFORMATION MATTERS MORA THAN PRIMARY REWARD

PATRICK ANSELME

2023

La teoría de forrajeo óptimo indica que los animales intentarán maximizar la recompensa por tiempo a largo plazo, pero describe agentes ideales con conocimiento completo del entorno, es decir, se centra en el problema de la explotación y no en la exploración. Sin embargo, la exploración es fundamental para la maximización.

La exploración puede verse como adaptación a las variables que interfieren con la maximización.

En algunas especies, como las hormigas y abejas, individuos separados hacen tareas de exploración y explotación. Esta disociación indica que son procesos separados que dependen de procesos bioconductuales diferentes.

Para algunos, la exploración es producto de la curiosidad, *i.e.*, aversión a la incertidumbre y su evitación mediante la colecta de información. Para otros, la exploración regula la activación hacia abajo cuando es alta, y hacia arriba cuando es baja. Aquí, se argumenta que la curiosidad no es requisito de la exploración, pues se ha encontrado en organismos sin cerebro. Más bien, la recompensa no garantizada es vista como un desafío por superar. La incertidumbre, aunque aversiva, no suele generar evitación. Los organismos evolucionaron para enfrentar desafíos en el ambiente, y esta reacción a la incertidumbre es benéfica para la salud y supervivencia.

La explotación es un fenómeno local relacionado con el consumo; la exploración es global, y pretende determinar dónde, cuándo y cómo encontrar recompensas o sus claves asociadas. La exploración es una inversión para el futuro. Sin embargo, si las claves, recompensas, o incluso el hambre, no motivan la exploración, ¿qué lo hace?

Tres principios se proponen: *consistency tracking*, los organismos expuestos a recompensa no garantizada buscan información sobre la consistencia de pareos clave-recompensa más que las claves y recompensas mismas; *incentive effort*, al estar expuestos a recompensa no garantizada, los organismos alargan la exploración para compensar la falta de control cognitivo que tienen en la situación; y *behavioral variability*, los organismos expuestos a recompensas no garantizadas amplían el rango espacial o temporal de su comportamiento dado que esto incrementa las posibilidades de nuevos encuentros exitosos.

1. Resource exploitation and the local influence of cues

La expectativa o detección de un estímulo apetitivo lleva a una aproximación no aleatoria a él. La saliencia incentiva (“wanting”) es el proceso que transforma la representación cerebral de una mera memoria o percepción a un incentivo motivacionalmente potente. El proceso depende de la liberación de dopamina en el estriado ventral y dorsal. Para algunos organismos en procedimientos pavlovianos el pareamiento entre comida y señal hace a la señal atractiva en sí misma (*sign-trackers*). Esto parece concordar con las nociones dichas: la procuración de recompensa es el factor a optimizar, pues una clave que predice mayor calidad o cantidad de comida debería ser aproximada y preferida sobre otras. Lo mismo sucede en procedimientos instrumentales: reforzadores de mayor calidad generan preferencia. Pero el argumento del artículo es que la influencia de asociaciones clave-consecuencia y acción-consecuencia es limitada, dado que son importantes al explotar locaciones bien conocidas, y por tanto no pueden explicar la conducta a una escala mayor.

Hay una distinción clásica entre búsqueda focal y general, sin embargo, si ambas buscan claves y recompensas, ¿qué distingue una de la otra? ¿Qué se busca en cada una? ¿Y cómo deciden los animales cambiar entre ellas? La saliencia incentiva y la ley de igualación pueden explicar la búsqueda focal, donde se aproxima un estímulo detectado o esperado, pero no es así con la búsqueda general, que no pretende aproximar un estímulo, sino contrarrestar los efectos de la incertidumbre mediante la exploración. La explicación de la exploración requiere una teoría enfocada en la información y no en la explotación inmediata de claves y recompensas.

2. Consistency tracking

Breves ocurrencias de explotación durante la exploración (consumir un ítem encontrado al explorar) no deben llevar a concluir que la exploración no es más que explotación a mayor escala. En la exploración los organismos intentan determinar *cuándo, dónde y cómo* encontrar asociaciones que predigan con la mayor consistencia posible la presencia de recompensas. Sin *consistency tracking* la supervivencia estaría sujeta solo al control de encuentros positivos y negativos.

Consistency tracking es adaptativo porque maximiza las buenas decisiones y minimiza las malas en un ambiente. Este principio explicaría por qué ocurre la búsqueda exploratoria en entornos nuevos en ausencia de expectativa de recompensa (donde se pueden encontrar posibles consistencias de clave-recompensa), pero no se muestra en entornos familiares donde la recompensa es esperada (y ya se conocen las consistencias).

Los procedimientos de respuesta de observación han mostrado que los animales valoran la consistencia de las señales y las consecuencias, pero quizá la evidencia más robusta venga del procedimiento de elección subóptima. Se han propuesto diversas explicaciones, como el contraste, la reducción de la demora, y otras, pero ninguna da una explicación completa de los hallazgos.

Un hallazgo indica que si las demoras de entrega son menores en la alternativa subóp-

tima que en la óptima, la elección subóptima es más probable. Es posible que los organismos hayan evolucionado para percibir una clave predictora después de una demora como más consistente, y que estén mostrando una preferencia por la consistencia de los pareos clave-comida (información) que por la comida misma (recompensa).

Un experimento con humanos comparó tres modelos de búsqueda de información no instrumental. El modelo con el mejor ajuste con los datos fue aquél que suponía a la incertidumbre como un estado aversivo, y los investigadores concluyeron que al buscar información no instrumental, las personas buscan la resolución de la incertidumbre más que disfrutar de la anticipación de consecuencias futuras.

3. Conditioning, higher-order conditioning, and occasion setting

Un *occasion setter* es un estímulo que indica si está activa una contingencia específica (un E^D). Este fenómeno puede contribuir a explicar la elección subóptima. El estímulo del eslabón inicial puede funcionar como un *occasion setter* que indica el contexto (la opción) que contiene información; en cierto sentido, se crea un contexto que resuelve la ambigüedad. La otra alternativa, en cambio, no indica la presencia de información.

4. Hoarding behavior as an example of consistency tracking

Guardar comida fuera del propio cuerpo para evitar que otros tengan acceso a ella es acumulación. La acumulación puede verse como un intento por disminuir la variabilidad en la disponibilidad de comida para minimizar el riesgo posterior de inanición.

5. Incentive effort

Esfuerzo incentivo significa que la incertidumbre hace al esfuerzo auto-motivador.

Se argumenta que la incertidumbre motiva la búsqueda porque es percibida como un desafío por resolver. Idea extraña, dado que implica que para los organismos es reforzante “resolver” un desafío.

En el procedimiento de PCA los animales responden más ante un estímulo que inconsistentemente predice comida que a uno consistente. Si el esfuerzo ante un desafío no es reforzante por sí mismo, ¿por qué los animales responden más bajo la incertidumbre? Sinceramente no me convence.

Una explicación simple sería que la atribución de saliencia incentiva es mayor para un estímulo inconsistente que para uno consistente. Sin embargo, se ha evaluado la saliencia incentiva en condiciones de certidumbre e incertidumbre y no se han encontrado diferencias, por lo que quizá no se puede atribuir a ello la diferencia en esfuerzo.

Cuando el entorno es incierto y el aprendizaje no ayuda, la única opción para sobreponerse a la aversión de la incertidumbre es trabajar más. Esto se mostró en un experimento con palomas en el que se colocó comida en una tabla con agujeros en dos condiciones: una de incertidumbre con 30 de 90 agujeros con comida; y otra de certidumbre con 30 de 30 agujeros con comida. Las palomas pasaban más tiempo y se esforzaban más por visita en la condición de incertidumbre. Me parece aventurado concluir que la inconsistencia hace que las palomas estén dispuestas a trabajar más, pero eso es lo que se concluye.

Aunque se piensa que la igualación es óptima, quizá en un entorno de incertidumbre la estrategia óptima para mantenerse con vida es la sobre-igualación (responder más de lo esperado con base en la tasa de reforzamiento por agujero).

6. Contrafreeloading as an example of incentive effort

Bajo ciertas circunstancias los animales prefieren comida ganada sobre comida gratuita. Por ejemplo, jerbos a los que se da un tazón con 1000 semillas libres y otro con 200 semillas escondidas en arena consumen más del segundo tazón, y esto ocurre aunque no estén hambrientos, por lo que la aproximación a la comida no parece ser la función principal.

Se propone que la presencia de incertidumbre es un factor que motiva el contrafreeloading: cuando se resuelve la incertidumbre la alternativa libre es preferida. Aunque esto tiene beneficios evolutivos a largo plazo (el *qué*), no indica los mecanismos proximales (el *cómo*).

La explicación del esfuerzo incentivo indica que los organismos invierten más tiempo y esfuerzo para compensar su falta de control cognitivo.

7. Behavioral variability

La incertidumbre resulta del mundo cambiante. Aun si tuviesen memorias perfectas, los animales solo tienen acceso a correlaciones imperfectas entre eventos.

La expectativa de reforzamiento correlaciona negativamente con la variabilidad conductual, es decir, al esperar omisión, los animales exploran fuentes alternativas.

Introducir mayor variabilidad en las respuestas puede parecer similar a ejercer un mayor esfuerzo en la tarea, pero en realidad variabilidad y esfuerzo son independientes. Además, la variabilidad es inversamente a la probabilidad p , mientras que el esfuerzo es proporcional a la incertidumbre u , expresada en términos de probabilidad como $p(1 - p)$. Esto implica que tiene sentido esforzarse cuando la recompensa es localmente incierta pero globalmente esperada; en contraste, la variabilidad incrementa la probabilidad de nuevas oportunidades en lugar de persistir en una estrategia inútil.

8. Relocation as a form of behavioral variability

Al encontrar escasez de comida los animales muestran movimiento de Lévy (en línea recta por una distancia relativamente grande y una dirección aleatoria), seguido de movimiento Browniano (pasos pequeños en direcciones aleatorias que los mantienen cerca de su lugar inicial).

Mostrar movimiento de Lévy puede ser adaptativo, y se sugiere que es un indicador de transición entre exploración y explotación.

9. A need to know or the perception of challenges?

Los tres principios descritos hacen posible la búsqueda de información sin una “necesidad de saber”, el cual es un concepto problemático.

Necesidad de saber implicaría que cualquier información es una recompensa, pero los organismos ignoran prácticamente toda la información que les rodea. El artículo defiende que la única información que es buscada activamente es la de relaciones entre clave y recompensa o acción y recompensa.

Además, la necesidad de saber es un concepto vago y enigmático, pobremente definido.

Se argumenta que enfrentarse a la incertidumbre es parte del bienestar de los animales, y que animales sin acceso a estimulación cognitiva enfermarán.