

PARADOXICAL CHOICE AND THE REINFORCING VALUE OF INFORMATION

VICTOR AJUWON ANDRÉS OJEDA ROBIN A. MURPHY
TIAGO MONTEIRO ALEX KACELNIK

2022

Bienes de poco acceso u otros resultados de beneficio sustancial son reforzadores efectivos. Recientemente se ha comenzado a cuestionar si, además de estos bienes, la información (reducción en la incertidumbre) modula el comportamiento mediante los mismos procesos que las recompensas convencionales, es decir, si la información puede ser un reforzador primario.

La información es un recurso valioso que puede usarse para mejorar la eficiencia del forrajeo, apareamiento, transporte, o más. Los animales pueden buscar información antes de hacer elecciones y esto puede mejorar la adquisición de bienes. El valor adaptativo y de reforzamiento de la búsqueda de información se derivan de su habilidad para incrementar un beneficio bien definido.

Lo que parece paradójico es que en campos como la micro economía, la teoría de forrajeo, y modelos clásicos de aprendizaje por refuerzo, es que los animales prefieren señales informativas en casos en que no tienen uso instrumental potencial, es decir, son adversos a la incertidumbre o “curiosos”.

La idea de que los animales valoran la información se postuló para explicar el fenómeno de las respuestas de observación, en el que los sujetos pueden resolver la incertidumbre sobre contingencias por venir al hacer una respuesta, aunque la información no puede usarse para modificar el resultado.

Han habido algunas hipótesis para explicar el fenómeno. Una de ellas, que será llamada hipótesis de “información” o de “reducción de la incertidumbre”, sugiere que la información es en sí misma intrínsecamente reforzante porque resuelve la incertidumbre, que tiene un valor hedónico negativo. Un mecanismo así podría haber evolucionado si la información es a menudo asociada con beneficios sustanciales en contextos ecológicos y no es muy costosa de adquirir.

Se ha sugerido que la información es valiosa porque previene una incertidumbre prolongada, que se asume como aversiva. También se ha propuesto que la información podría derivar su valor reforzante de permitirle a los sujetos anticipar positivamente (*savour*) los resultados mientras los esperan.

Una explicación alternativa, que será nombrada hipótesis de reforzamiento condicionado, indica que la señal para buenas noticias S^+ en tareas de respuestas de observación adquiere propiedades de reforzamiento secundarias dado que está pareada con comida y se convierte en un reforzador condicionado.

Se ha argumentado que es el S^+ , una vez asociado con la comida, lo que refuerza la conducta en la tarea de respuestas de observación. La dificultad es que por el mismo razonamiento la señal de malas noticias debería ser un condicionador secundario para la ausencia de consecuencia. La explicación se sostiene solo si el S^+ incrementa la respuesta más de lo que S^- la disminuye.

Estos mecanismos no son mutuamente excluyentes. Uno postula que la información es reforzante *per se*, mientras el otro que las señales para comida pueden adquirir propiedades de reforzamiento secundarias debido a su contingencia con la comida. Ambos mecanismos son biológicamente plausibles.

Para decidir entre las hipótesis se han hecho experimentos en que la comida o su ausencia no son anteceditos por una señal, es decir, se omite la presentación de S^+ o S^- . De acuerdo con la hipótesis de información la preferencia por la opción informativa debería adquirirse y mantenerse por S^+ o S^- indistintamente. De acuerdo con la hipótesis de reforzamiento condicionado solo debería adquirirse o mantenerse cuando la señal de buenas noticias S^+ está presente. Estos experimentos han encontrado que S^- solo no es suficiente para mantener las respuestas de observación. Estos resultados indican que la ganancia de información no es suficiente para explicar las preferencias observadas.

Resultados de protocolos similares al de respuestas de observación han reavivado el interés en la posibilidad de que la información sea intrínsecamente reforzante. Monos parecen preferir recibir señales no ambiguas sobre la magnitud de la recompensa venidera, y están dispuestos a renunciar a reforzamiento por obtenerlas. Estas preferencias correlacionan con actividad en neuronas implicadas en la representación de ganancias primarias, lo que sugiere un valor intrínseco para la información.

Experimentos de elección paradójica han encontrado que palomas, estorninos, y ratas prefieren una alternativa con información que no pueden utilizar incluso cuando la alternativa con información ofrece menos recompensas. El hecho de que los animales renuncien a reforzamiento primario para generar señales predictivas aparentemente sin función es buena razón para sospechar un valor de reforzamiento primario en la reducción de la incertidumbre.

Se realizaron experimentos en elección paradójica en ratas manipulando la saliencia de las señales predictivas en tres condiciones. Había dos opciones con redituabilidad idéntica. Un tiempo fijo tras la elección las alternativas entregaban comida con 50 % de probabilidad. En la opción informativa el resultado era señalado entre la elección y el resultado; en la otra, no. Dado que la señal ocurría después de la elección, la información no podía usarse para modificar la probabilidad de obtener comida.

En el tratamiento S^+S^- los estímulos discriminativos eran sonidos diferenciables. En *solo_* S^- se omitía la señal positiva, y lo contrario ocurría en *solo_* S^+ .

La hipótesis de información predice que la clave positiva y negativa por sí solas deberían poder generar respuestas de observación. La hipótesis de reforzamiento condicionado indica que solo el estímulo positivo debería ser reforzante, y la presencia del negativo

debería reducir más que incrementar la adquisición de la respuesta de observación.

Si la preferencia por la alternativa informativa se sostuviera solo cuando se omite la señal negativa pero no la positiva, entonces habría soporte experimental para la hipótesis del reforzamiento condicionado.

Se registrarán dos medidas de preferencia: la proporción de elecciones en ensayos de dos opciones y las latencias en ensayos forzados.

1. Methods

El experimento se realizará en cámaras con tres palancas: dos en el panel frontal a ambos lados del comedero, y una en el trasero, colocada en el centro.

2. Training

Después de habituación y entrenamiento a palancas, se realizó un protocolo pavloviano en el que las ratas fueron expuestas a las claves y sus contingencias respectivas.

3. Experimental procedures

Todos los ensayos comenzaban con la presentación de la palanca trasera. Presionarla la retraía y extendía una o ambas palancas frontales. Presionar las frontales iniciaba su eslabón terminal asociado con un estímulo dependiente de la condición y la entrega y omisión de reforzamiento. En la alternativa informativa los estímulos indicaban confiablemente la consecuencia final; en la no-informativa, no.

En la condición S^+S^- los eslabones terminales positivos y negativos eran señalados por claves diferenciables. En *solo_* S^+ se sustituía la clave negativa con un blackout, y en *solo_* S^- sucedía lo mismo con la positiva.

4. Data analysis

Para medir la preferencia con base en la latencia, para cada individuo se calculó un índice llamado $L_{(Info)}$ de la siguiente forma:

$$L_{(Info)} = \frac{R_{(Info)}}{R_{Info} + R_{NoInfo}}$$

donde R_{Info} y R_{NoInfo} son las medianas de latencia para responder a la alternativa informativa y no informativa, respectivamente. Valores mayores a .5 indicarían preferencia por la alternativa informativa y viceversa.

5. Results

Para garantizar que los animales podían discriminar las contingencias se determinó el tiempo total que pasaban con la cabeza dentro del comedero en presencia de cada una de las señales. Se encontró que pasaban mayor tiempo ante S^+ , intermedio en los estímulos no discriminativos, y menor en S^- .

En ensayos de elección se encontró preferencia por la alternativa informativa en todas las condiciones, con la única diferencia de que la adquisición fue más lenta en *solo_ S^-* . Se encontraron efectos significativos de tratamiento, sesión, e interacción, lo que refleja la adquisición más lenta en S^- .

Se utilizó la mediana de latencias para cada alternativa para el análisis. En todas las condiciones las latencias eran menores para la alternativa informativa que para la no informativa en las últimas sesiones. Aunque la latencia para la alternativa informativa era relativamente constante, la latencia para la no informativa varió: era larga en S^+S^- , intermedia en *solo_ S^-* , y corta en *solo_ S^+* , a pesar de que la alternativa no informativa era idéntica entre tratamientos.

El índice L_{Info} fue similar en *solo_ S^+* y *solo_ S^-* , pero en ambos grupos fue mayor que en S^+S^- . El índice estuvo por debajo de .5 para todos los tratamientos, lo que es consistente con preferencia por la alternativa informativa. En resumen, la preferencia era más fuerte cuando ambas consecuencias eran señaladas explícitamente, e igualmente fuerte si alguna de las señales era omitida.

Las entradas al comedero reflejaban buena discriminación. El nivel absoluto de entradas parecía inversamente relacionado a cuánta señalización estaba disponible, como si la atención a las señales explícitas compitiera con la investigación exploratoria del dispensador de comida.

6. Discussion

Para comenzar, nombrar “subóptima” a la preferencia es inexacto dado que implica que la conducta es desadaptativa en contextos ecológicos. “Paradójica” es mejor dado que se refiere solo al punto de vista del observador.

Al presentar dos opciones que difieren solo en la predictabilidad post-elección, las ratas, al igual que aves y primates, prefieren fuertemente la alternativa informativa.

Estudios previos no han encontrado esta misma preferencia, aunque en ellos probablemente se debió a que la alternativa informativa tenía una menor tasa de reforzamiento.

La preferencia asintótica por la alternativa informativa es robusta a pesar de la ausencia de una señal saliente de buenas o malas noticias. Este hallazgo es consistente con observaciones similares en estorninos, monos y humanos. Aunque parece que la ausencia de una señal positiva saliente ralentiza la adquisición de la preferencia por la alternativa informativa, mientras que la ausencia de una señal negativa tiene un efecto leve de acelerar la misma adquisición. Esto es consistente con estudios que muestran que en ratas el estímulo negativo adquiere propiedades inhibitorias.

El análisis de latencias se basa en los supuestos del modelo de elección secuencial, que

postula que la elección puede verse como una carrera entre dos procesos paralelos pero independientes de valoración de las alternativas. Medir la conducta mediante más de un procedimiento es importante dado que, si los fenómenos son robustos, se debería encontrar invarianza procedimental.

El análisis de latencias mostró que las variaciones entre tratamientos eran mediadas solo por diferencias en latencias en la alternativa no informativa, a pesar de que su programación era idéntica entre condiciones. Es decir, los efectos de condición eran mediados por el cambio en latencias en la alternativa menos preferida.

La hipótesis de información indica que los animales son adversos a la incertidumbre, mientras que la de reforzamiento condicionado dice que la preferencia por la alternativa informativa incrementa debido a las señales positivas, y decrementa con las negativas, y la influencia excitatoria es vista como mayor que la inhibitoria.

Ambos mecanismos tienen dificultades: la mera adquisición de información no da a los sujetos la habilidad de incrementar las recompensas, aunque esto puede abordarse argumentando que, en la naturaleza, la información sobre bienes relevantes es a menudo útil hasta el punto de que la evolución puede diseñar funciones de utilidad que pueden ser engañadas por los protocolos experimentales. Por ejemplo, la información de eventos desfavorables permite a los animales abortar una persecución. Además, aun si la información no se puede usar de inmediato, puede ayudar a resolver problemas futuros. Así, sería la artificialidad de no poder usar la información en el protocolo lo que genera la paradoja.

Sobre la hipótesis de reforzamiento condicionado, aunque no hay razones a priori sobre por qué el efecto excitatorio de S^+ debería ser mayor que el inhibitorio de S^- , es probable que en la naturaleza las claves que señalan la presencia de bienes son más prevalentes y confiables que las que señalan su ausencia. Los poderes excitatorio e inhibitorio no necesitan ser simétricos.

Estos resultados no apoyan decididamente a ninguna de las hipótesis, y ambos mecanismos podrían actuar simultáneamente.

La hipótesis de la información hace dos predicciones únicas: que S^- por si solo debería mantener la preferencia por la alternativa informativa, lo que de hecho es apoyado por los datos. Aunque los datos de entrada al comedero parecen sugerir lo contrario: cuando había una clave explícita S^- las entradas eran menores que ante S^+ o claves no informativas, indicando propiedades inhibitorias para S^- . Pero aunque puede inhibir las entradas a comedero eso no implica que no pueda reforzar la elección por la alternativa informativa, lo que ocurre más temprano en el ensayo, mediante la reducción de la incertidumbre.

Otra predicción es que S^+ y S^- deberían reforzar en la misma medida. El hallazgo de que la omisión de S^+ ralentiza la adquisición más que la omisión de S^- es incongruente con esta predicción.

Entre las explicaciones sobre cómo S^+ adquiere valor como reforzador condicionado están la hipótesis del contraste, la *stimulus value hypothesis*, *Signals for Good News*, el *temporal information model*, y la *selective engagement hypothesis*. Todas comparten el supuesto de que S^+ es responsable de la preferencia, y que el S^- tiene poca o nula influencia.

Podría argumentarse que la manipulación no eliminó el reforzamiento condicionado de la alternativa informativa: quizá el compuesto de la respuesta con la ausencia de una

clave fue tomada como un estímulo condicionado en sí mismo. Así, habría reforzamiento condicionado aun sin una clave explícita.

Esto podría explicar la preferencia y velocidad de la adquisición si se pareara con el fenómeno del *feature-positive effect*, en el cual el aprendizaje en discriminación que involucra la presencia o ausencia de una característica, la presencia es más fácil de asociar con un resultado positivo que la ausencia. Eso explicaría por qué la preferencia por la alternativa informativa se desarrolló más rápido en *solo_ S^+* que en *solo_ S^-* .

Los resultados son también consistentes con la posibilidad de la reducción de la incertidumbre y el reforzamiento condicionado actúen en conjunto para generar la preferencia paradójica. La preferencia por la alternativa informativa en ausencia de un S^+ explícito es consistente con la predicción de que un S^- perceptible es suficiente para generar preferencia por la alternativa informativa mediante la reducción de la incertidumbre. La adquisición más rápida en *solo_ S^+* comparada con *solo_ S^-* apoya la presunción de la hipótesis de reforzamiento condicionado de que un S^+ perceptible refuerza la elección de la alternativa informativa. Este mecanismo capturaría dos bienes relevantes—la cantidad de información y su contenido (buenas o malas noticias)—como factores que moldean la adquisición de preferencias en elección paradójica.

En resumen, las ratas muestran preferencia robusta por información no-instrumental y esta preferencia es más fuertemente influenciada por buenas que por malas noticias. Los efectos de tratamiento fueron mediados por diferencias en latencias para la alternativa menos preferida. Aunque la reducción de la incertidumbre probablemente no sea la única explicación para la preferencia por la información anticipada, la evidencia indica que tiene un papel importante junto con el reforzamiento condicionado.