On the value of advanced information about delayed rewards

Alejandro Macias Armando Machado
Marco Vasconcelos

2023

En ciertas preparaciones los animales, aunque hambrientos, renuncian a comida a cambio de información inútil. Sin embargo, aquí se dice que las ratas también prefieren la alternativa informativa y se cita a Ajuwon, Chow, y Cunningham y Shahan.

La preferencia se parece a la hipótesis de búsqueda de información propuesta en la literatura de respuestas de observación. Originalmente, se presentaba a palomas una tecla blanca donde se alternaban dos programas de reforzamiento equiprobables, no señalados, e impredecibles. Uno ofrecía comida cada 30 s; el otro, extinción. Si la paloma pisaba un pedal la tecla se encendía en verde o rojo dependiendo del programa vigente. Aunque la presión del pedal no podía cambiar el resultado las palomas aprendieron a presionarlo. De modo similar, ratas prefieren alternativas que informan sobre un choque inescapable por encima de alternativas sin información.

Si esta tendencia a preferir la información es general debería extenderse a información sobre cualquier evento biológicamente relevante, sea su cantidad, demora, probabilidad o calidad.

Bromberg-Martin mostró que macacos prefieren conocer la magnitud del reforzamiento que recibirán aun si no pueden alterarlo.

En este experimento se evalúa si los animales preferirán conocer anticipadamente la demora de la comida cuando ambas alternativas tienen la misma densidad de reforzamiento.

Una modificación del modelo Δ-Σ indica que los animales deberían preferir las recompensas señaladas, y que además la magnitud de la preferencia debería depender de la disparidad entre las demoras posibles.

Según del modelo, en elección subóptima la preferencia depende de Δ, la diferencia entre probabilidades de reforzamiento anunciadas por los estímulos de cada opción; y Σ, la probabilidad global de reforzamiento de cada opción. El valor de las opciones es dado por

$$V_i = (\Sigma_i)^c * e^{\beta * \Delta_i},$$

donde los parámetros de escalamiento c y β son mayores que 0, y i representa la opción informativa o no informativa. La preferencia puede estimarse con la razón de Luce

$\frac{V_{\text{info}}}{V_{\text{info}}+V_{\text{no-info}}}$, que se simplifica a

$$P_{info} = \frac{1}{1 + \left(\frac{\Sigma_{non-info}}{\Sigma_{info}}\right)^{c} e^{-\beta(\Sigma_{info}-\Sigma_{non-info})}}.$$

Sin embargo, esta ecuación predice indiferencia dado que las probabilidades de reforzamiento son iguales (así que Δ_{info} y $\Delta_{\text{no-info}}$ son cero), y Σ es 1 para ambas. Sin embargo, aunque no hay contraste en probabilidad, sí lo hay en demora: supóngase una demora de 5 s y una de 20 s. El Δ de las inmediateces de la alternativa informativa sería $\frac{1}{5} - \frac{1}{20} = .15$, y el Δ de la alternativa no informativa sería 0 dado que el tiempo promedio es 12.5 ante cualquiera de los dos estímulos.

Dada esta fuente de contraste y dado que Σ es igual para ambas alternativas, la ecuación se reduce a una de un parámetro:

$$P_{\text{info}} = \frac{1}{1 + e^{-\beta\left(\frac{1}{d_s}-\frac{1}{d_l}\right)}},$$

donde d_s y d_l son las demoras corta y larga. Si $\frac{d_l}{d_s} = r$ y $d_s + d_l = S$ la ecuación se vuelve

$$P_{\text{info}} = \frac{1}{1 + e^{-\frac{\beta}{S}\left(r - \frac{1}{r}\right)}}.$$

Este modelo predice preferencia por la alternativa informativa cuanto más grandes sean las diferencias entre las demoras dentro de la alternativa; y predice indiferencia si la diferencia es de cero.

Se especula de un razonamiento ecológico según el cual sería ventajoso conocer las posibles demoras de captura de tipos de presa, por lo cuál podría existir un sesgo inherente en los animales para preferir alternativas señaladas. Según este razonamiento, el beneficio de preferir la información también incrementa según son mayores las diferencias entre las demoras, pues rechazar una presa cuando su demora de captura es alta con respecto a otra presa con demora menor incrementa en mayor medida la tasa media de captura.

Algunos argumentan que la información es reforzante *per se*, independiente de cualquier valor instrumental (*hipótesis del valor intrínseco*). Algunos hallazgos indican que el valor subjetivo de la información no instrumental comparte un código neural con las recompensas primarias (Bromberg-Martin & Hikosaka, 2009). Sin embargo, esta propuesta solo predice la preferencia por las demoras señaladas, pero dado que el valor de la información sería independiente de las diferencias en las demoras señaladas, no tendría que haber modulación por esta diferencia.

Se ha encontrado evidencia de preferencia por demoras señaladas cuando las diferencias entre demoras son grandes, pero el efecto de la similitud entre demoras no se ha estudiado.

Este experimento busca contrastar las hipótesis Δ-Σ, ecológica, y de valor intrínseco. La última difiere de las primeras solo en la falta de predicción de un efecto modulador para las diferencias entre demoras.

1. Method

1.1. Subjects

Siete palomas con experiencia en elección subóptima (Fortes et al, 2016).

1.2. Apparatus

Tres cajas con un panel de respuesta con tres teclas, comedero, y luz general.

1.3. Procedure

Preentrenamiento. Para evitar acarreo se entrenaron respuestas ante los colores y símbolos en diferentes programas FR.

Tarea experimental. Cada sesión constaba de 96 ensayos: 32 de elección y 64 forzados. En el eslabón inicial las teclas izquierda y derecha se iluminaban con una cruz y un círculo. Una respuesta en cualquiera apagaba ambas y llevaba a su eslabón terminal. Al elegirse la alternativa informativa las demoras corta y larga estaban señaladas por un color específico en la tecla central; al elegirse la no informativa, las demoras no estaban correlacionadas con el color de la tecla central. La localización de las alternativas cambiaba aleatoriamente tanto en ensayos forzados como de elección.

Cada paloma pasó por dos condiciones de línea base y tres experimentales. En línea base todas las demoras se ajustaron a 12.5 s; en las experimentales se incrementaron las diferencias entre las demoras, pero se mantuvo constante el promedio (10 vs 15, 7.5 vs 17.5, 5 vs 20). El orden de las condiciones fue contrabalanceado entre palomas. Cada condición duró al menos 10 sesiones y permaneció hasta obtener estabilidad.

2. Results

Las palomas tardaron 16 sesiones en promedio en alcanzar la estabilidad por condición, y el tiempo para llegar a estabilidad no varió entre condiciones.

Las palomas eran indiferentes con la razón long/short de 1.0, pero se encontró una preferencia por la información en las demás condiciones que incrementó en función de la diferencia entre demoras.

Las predicciones del modelo Δ-Σ modificado describen bien los datos.

Se ha mostrado que la latencia es una medida sensible de preferencia, así que se analizó la latencia a responder en los eslabones iniciales en ensayos forzados. Las latencias en línea base eran similares, así que se promediaron en una sola medida. Según incrementaron las diferencias, las latencias comenzaron a divergir y los animales mostraron latencias mayores para la alternativa menos preferida. Las latencias son consistentes con los datos de preferencia.

Finalmente se analizaron las tasas de respuesta a los estímulos de los eslabones terminales. Cuatro de siete palomas no picaban ninguno de los estímulos. Para el resto, las respuestas entre los estímulos no discriminativos permanecieron indistinguibles, y las

respuestas al estímulo que señalaba la demora corta en la alternativa informativa fueron mayores que para el estímulo complementario. Esto sugiere que esas palomas aprendieron las demoras señaladas por cada estímulo.