

# TIME, UNCERTAINTY, AND SUBOPTIMAL CHOICE

ALEJANDRO MACÍAS

VALERIA V. GONZÁLEZ

ARMANDO MACHADO

MARCO VASCONCELOS

2024

Una señal que quita ambigüedad sobre el resultado del forrajeo debería ser valiosa. En elección subóptima es así hasta el punto que los animales renuncian a comida para recibir información.

La búsqueda de un mecanismo ha llevado a proponer que la elección subóptima es un efecto del valor de reforzamiento condicionado diferencial de los estímulos de los eslabones terminales: el estímulo que señala reforzamiento seguro adquiere propiedades reforzantes por su relación predictiva con la comida. Sin embargo, no se ha resuelto por qué este efecto sería superior al efecto del estímulo inhibitor. Aun así, se ha mostrado que la confiabilidad de la relación predictiva es crucial.

En este experimento se explora el efecto no de manipular las probabilidades anunciadas por los estímulos, sino de cambiar el tiempo bajo incertidumbre después de la elección. Al elegir la opción informativa se presenta el estímulo  $S^G$  durante los primeros  $t$  segundos de la demora, seguido de lo cual el color puede mantenerse en  $S^G$  o cambiar a  $S^R$ . La incertidumbre se prolonga entonces durante  $t$  segundos. La evidencia indica que posponer la resolución de la ambigüedad debería disminuir la preferencia por la alternativa informativa.

En el segundo experimento se hizo la manipulación complementaria: ahora se encendía  $S^R$  por default y cambiaba o se mantenía en su color tras  $t$  segundos.

Se varió  $t$  de 0 a 10 s. Se espera que cuando  $t = 0$  el resultado sea el de elección subóptima normal; y cuando  $t = 10$  la preferencia se revierta, dado que la alternativa pasa a ser no informativa. Se predice que la preferencia por la opción informativa disminuirá monótonicamente con  $t$ .

## 1. Experiment 1 - Temporarily reducing the reliability of $S^G$

Todo ensayo informativo empezaba presentando  $S^G$ , que podía cambiar o mantenerse después de  $t$  segundos.

Los sujetos fueron 7 palomas en cajas individuales. Se utilizaron cajas operantes con un panel de respuesta con tres teclas circulares. Bajo la tecla central había un dispensador de comida, y en la pared opuesta había una luz general.

## 1.1. Pretraining

Las palomas fueron entrenadas con programas de FR con los seis estímulos del experimento (rojo, verde, amarillo, azul, + y ○).

## 1.2. Training

Cada paloma pasó por 9 longitudes de  $t$ , cada una de al menos 12 sesiones de duración y hasta alcanzar estabilidad.

Cada sesión consistía en 120 ensayos, 40 de elección y 80 forzados, con las condiciones normales de elección subóptima más la manipulación ya descrita.

Todas las palomas comenzaron con  $t = 0$  (procedimiento estándar), seguido de  $t = 5$ . Después, cuatro palomas pasaron por  $t = 1.5 \rightarrow t = 8.5$ , mientras tres más pasaron por  $t = 8.5 \rightarrow t = 1.5$ . Tras un regreso a línea base, todas pasaron por  $t = 3.25 \rightarrow t = 0.75 \rightarrow t = 6.75$  o el orden contrario. Finalmente todas experimentaron  $t = 10$ , momento en el cual la opción ya no era informativa.

## 1.3. Results and discussion

Se analizaron los resultados de las últimas tres sesiones de cada condición. Dado que no hubo diferencias entre las dos condiciones de  $t = 0$ , sus resultados se promediaron.

Como se esperaba, cuando  $t = 0$  hubo preferencia por la alternativa informativa. La preferencia por la alternativa informativa disminuyó con incrementos en  $t$ . La tasa de respuestas a la tecla incrementó según se aproximaba  $t$ . Después de  $t$  segundos las palomas continuaban picando la tecla si  $S^G$  prevalecía, pero dejaban de picar si  $S^R$  era presentado, con excepción de la condición de  $t = 0$ , donde las palomas picaban durante  $S^R$  infrecuentemente.

En cuanto a la alternativa no informativa, la tasa de respuesta fue consistentemente baja.

En resumen, la preferencia por la alternativa informativa disminuyó cuando el estímulo presentado no eliminaba la ambigüedad inmediatamente después de la elección.

## 2. Experiment 2 - Temporarily reducing the reliability of $S^R$

Todo ensayo informativo comenzaba con  $S^R$ , que podía cambiar a  $S^G$  después de  $t$  segundos o permanecer igual. Se hipotetizaba que el valor predictivo de la alternativa informativa disminuiría al no resolver la ambigüedad inmediatamente.

### 2.1. Procedure

Todos los detalles fueron idénticos al experimento 1.

### 2.2. Results and discussion

De nuevo se promediaron los resultados de las condiciones de  $t = 0$  al no haber diferencias entre ellos.

Cuando  $t = 0$  las palomas mostraron preferencia marcada por la alternativa informativa. La preferencia tendió a disminuir con  $t$ , aunque el efecto fue menos pronunciado que en el experimento 1. Una correlación rank-order de Spearman mostró que la tendencia era significativa para cuatro de seis palomas.

Prácticamente no hubo respuestas durante  $S^R$  tanto antes como después de  $t$ . Las respuestas comenzaron cuando se presentó  $S^G$ . En este caso la conducta estaba bajo control de los estímulos solamente, y no bajo control temporal como en el experimento 1.

En resumen, la preferencia por la alternativa informativa disminuyó con incrementos en  $t$ , pero el efecto no fue tan pronunciado como en el experimento 1. Hubo mayor variabilidad, y algunas palomas solamente revirtieron su preferencia cuando la alternativa informativa se volvió completamente no informativa ( $t = 10$ ).

### 3. General discussion

Se muestra que la preferencia paradójica por la información costosa no depende solamente en la información transmitida, sino también de cuándo se entrega esa información. Se encontró que inicialmente hay una preferencia fuerte por la alternativa informativa, pero esta preferencia disminuye según se demora la información. Se argumenta que la preferencia por la alternativa informativa incrementa con el tiempo en certidumbre, y disminuye con el tiempo en incertidumbre.

Idea: se podrían manipular independientemente las duraciones de los tiempos bajo certidumbre e incertidumbre para dar soporte a la hipótesis de la información.

Estos hallazgos son consistentes con un reporte previo (McDevitt et al, 1997), en el cual se introdujo un *gap* de 5 s antes de la presentación de los estímulos manteniendo constante la duración del eslabón terminal. Hubo pocos cambios en la preferencia cuando el *gap* precedía solamente a  $S^-$ , pero hubo un cambio sustancial cuando precedía solamente a  $S^+$ , a ambos estímulos informativos, o a todos los estímulos. Una diferencia entre los procedimientos es que McDevitt utilizó un estímulo neutro (la tecla apagada) para llenar el *gap*, mientras que aquí se utilizaron  $S^+$  y  $S^-$ . Solo el milagro sabe si esto tiene un efecto importante en los resultados.

Stagner et al (2015) usaron un procedimiento en el cual  $S^+$  cambiaba a  $S^-$  después de 2, 5 u 8 segundos en un eslabón terminal de 10 s. Encontraron que cambiar  $S^+$  por  $S^-$  bajaba la preferencia por la alternativa informativa con algo de evidencia que indica que  $t$  más largos llevan a preferencias más débiles.

Bromberg-Martin usó un procedimiento de elección en monos en el cual ambas alternativas eran informativas, pero diferían en la demora con que entregaban información. Los monos preferían la alternativa con la menor demora, siendo todo lo demás igual.

Se puede interpretar este procedimiento suponiendo que el eslabón inicial persiste hasta el evento que resuelve la ambigüedad a los  $t$  segundos. Esto sería consistente con investigación previa que manipula la duración de los eslabones mostrando que eslabones iniciales más largos disminuyen la elección subóptima. En general, la preferencia subóptima parece depender de eslabones iniciales cortos y eslabones terminales largos.

En ambos experimentos de este estudio hubo un decremento en la preferencia por la alter-

nativa informativa en función de  $t$ , pero el efecto fue más pronunciado para el experimento 1. Es decir, aunque en ambos casos la duración de certidumbre e incertidumbre fue la misma, la reducción en la preferencia fue mayor cuando la incertidumbre estaba señalada por el estímulo positivo que por el negativo.

Varias asimetrías pueden explicar las diferencias:

- El nivel de degradación de contingencias: en el experimento 1  $S^G$  engañó a los sujetos en el 80 % de los ensayos, mientras que en el 2  $S^R$  los engañó solo el 20 % de las ocasiones. Formalmente, cuando  $t = 0$  en ambos experimentos  $p(\text{reinforcer} | S^G) = 1.0$ , y  $p(\text{reinforcer} | S^R) = 0$ , lo que da una contingencia positiva perfecta. Pero cuando  $t > 0$ , en el experimento 1,  $p(\text{reinforcer} | S^G) = 0.2$ , y  $p(\text{reinforcer} | S^R) = 0$ , lo que resulta en una contingencia positiva débil de 0.2; mientras que en el experimento 2  $p(\text{reinforcer} | S^G) = 1$ , y  $p(\text{reinforcer} | S^R) = 0.2$ , lo que resulta en una fuerte contingencia positiva de 0.8. Es decir, la degradación de la contingencia es mayor en el experimento 1 (de 1.0 a 0.2) que en el 2 (de 1.0 a 0.8), por lo que la incertidumbre del resultado era mayor en el experimento 1.
- Según incrementaba  $t$ ,  $S^G$  señalaba la recompensa con una demora promedio mayor en el experimento 1. Por ejemplo, cuando  $t = 0$ ,  $S^G$  señala comida tras 10 s; cuando  $t = 1.5$ , señala comida tras  $10 + (4 \times 1.5)$  segundos (los 10 s de cada ensayo reforzado más 1.5 s de los ensayos no reforzados que son 4 veces más frecuentes); y cuando  $t = 8.5$ , la demora promedio incrementaba a 44 s. En contraste, en el experimento 2 los incrementos en  $t$  hacían que  $S^G$  señalara la comida con demoras cada vez menores, iguales a  $10 - t$ . Esta asimetría podría haber desviado la preferencia.
- Los ensayos reforzados en el experimento 2 implicaban un cambio exteroceptivo de  $S^R$  a  $S^G$ , lo que implica que las palomas podrían involucrarse en otras actividades durante  $S^R$ , y retornar a picar cuando había un cambio a  $S^G$ , por lo que la espera podría implicar un costo de oportunidad menor en el experimento 2 que en el 1.

Una explicación de las diferencias requeriría de un modelo que relacione la confiabilidad de las señales con la probabilidad y demora, y que especifique su interacción con el costo de oportunidad.

Estos hallazgos junto a los de McDevitt y Stagner presentan el desafío de integrar la probabilidad y la demora en los modelos de elección subóptima. La hipótesis  $\delta - \Sigma$ , por ejemplo, hace predicciones utilizando las probabilidades de reforzamiento de los eslabones terminales, pero no tiene manera de explicar los hallazgos actuales, pues no integra en ningún sitio la duración de los eslabones. Una posibilidad es hacer que la razón entre las duraciones de los eslabones terminales amplifique los efectos de la diferencia en probabilidades.

Estos resultados señalan la complejidad de elección subóptima, y presentan el desafío de integrar los parámetros temporales con los probabilísticos.