Suboptimal choice: a review and quantification of the signal for good news (SiGN) model

Roger M. Dunn Jeffrey M. Pisklak
Margaret A. McDevitt Marcia L. Spetch

2023

Aunque la elección subóptima puede venir de procesos adaptativos en su entorno natural, disminuye la cantidad total de alimento obtenida. Es inconsistente con la ley del efecto, aprendizaje por refuerzo, y teoría de forrajeo óptimo. Pero el estudio de conducta aparentemente subóptima puede permitir entender las variables que controlan la conducta.

Los primeros reportes de conducta subóptima mostraron que se podía propiciar cuando el eslabón inicial requiere una sola respuesta, cuando hay gran demora entre la respuesta y la consecuencia (TL largos), y especialmente cuando la alternativa subóptima es informativa.

Un intento temprano para explicar la elección subóptima fue el modelo de Signal for Good News. Sus supuestos son que en programas de cadenas concurrentes probabilísticas la elección es dirigida por las diferencias en las tasas de reforzamiento primario (comida) y secundario (entrada en los eslabones terminales de la cadena); que un evento es un reforzador condicionado en función de la reducción de la demora a la comida;que los estímulos que predicen una reducción en la demora más allá por la señalada por la propia entrada al eslabón terminal son buenos reforzadores condicionados; y que las demoras señaladas a la omisión de comida tienen como única función crear un contexto de incertidumbre. El modelo da cuenta cualitativamente de los efectos de los parámetros temporales de la tarea y la relevancia de las señales.

Otras explicaciones han aparecido: decaimiento hiperbólico, contraste, consideraciones de forrajeo, información temporal, diferencias en probabilidades de reforzamiento en eslabones terminales, decaimiento de asociabilidad, persecución de errores de predicción anticipados y búsqueda de información, y mecanismos evolutivos que enfatizan las señales confiables de comida. Algunos modelos han sido formalizados matemáticamente. Curiosamente, SiGN no lo ha sido a pesar de ser de las explicaciones más antiguas. Aquí se intenta formalizar.

1. Suboptimal concurrent-chains procedure and literature review

1.1. Signaled 50 % versus 100 % food outcomes

Experimentos que comparan 50 % señalado contra 100 % suelen tener mucha variabilidad entre sujetos, lo que lleva a sospechar que las palomas son indiferentes y solo eligen debido a sesgos. Aunque aun la indiferencia entre las dos alternativas es en sí misma subóptima, la diferencia entre indiferencia y preferencia por la alternativa subóptima tiene implicaciones importantes para el modelamiento.

Los hallazgos generales de experimentos con estas probabilidades, con una sola respuesta como requisito, y con longitudes iguales de eslabón terminal, indican que cuando las demoras son cortas (menos de 10s) las palomas tienden a preferir la alternativa óptima; pero al alargarlas incrementa la elección por la alternativa señalada.

1.2. Signaled 20 % versus 50 % food outcomes

Estos estudios tienen eslabones iniciales de FR 1; y terminales de FT 10s. Se ha encontrado una preferencia consistente por la alternativa subóptima.

1.3. Role of signals

Correlacionar los estímulos terminales con la comida en la alternativa subóptima incrementa la preferencia de las palomas por ella. Si no existe correlación, no se desarrolla la preferencia subóptima.

También importa la contigüidad entre la respuesta y la presentación de la señal: al introducir un *gap* de 5 s entre la respuesta y el inicio del estímulo disminuyó sustancialmente la preferencia por la alternativa subóptima. Además, cuando el *gap* se inserta antes del encendido de la señal S^+ en la alternativa de 50 %, disminuyó en gran medida la elección subóptima, lo que no ocurrió si el *gap* se insertaba antes de la señal S^-, o antes de la señal de la alternativa de 100 %. Esto sugiere que el encendido de la señal de comida en un contexto de incertidumbre es importante para la elección subóptima, pero la señal no discriminativa no lo es.

1.4. Role of partial signals

Aproximaciones recientes intentan determinar el efecto de alternativas parcialmente señaladas. Fortes mostró que agregar entregas de comida después de algunas de las presentaciones de S^- tenía poco efecto en la elección, pero parecía devaluar la alternativa subóptima al evaluar con un procedimiento de ajuste de demora. González *et al.* manipularon la probabilidad de comida tras un estímulo específico manteniendo la probabilidad global de las alternativas constante, y encontraron que el grado en que los distintos estímulos predicen la comida se correlaciona directamente con la preferencia. Sears *et al.* aisló los roles de S^+ y S^- igualando la probabilidad de reforzamiento primario y proveyendo

una única señal "pura" para algunas alternativas, *e.g.,* la alternativa de "buenas noticias" llevaba a una señal de comida segura o una señal incierta;'la alternativa de "malas noticias" llevaba a una señal de omisión segura o una señal incierta; la alternativa no señalada llevaba siempre a estímulos inciertos. Las tres alternativas tenían la misma tasa de reforzamiento primario, pero la preferencia fue "buenas noticias" >"malas noticias" >"no señalada".

2. The role of temporal variables

Se ha manipulado la duración de los eslabones iniciales y terminales, y el intervalo entre ensayos.

2.1. Terminal-Link duration and schedule

Spetch *et al.* encontraron que eslabones terminales cortos (5 - 10s) llevaban a baja elección subóptima, pero eslabones largos la incrementaban. McDevitt encontró mayor elección subóptima con eslabones de 20s que con eslabones de 5s.

Un problema con el estudio de eslabones terminales es que el requisito de respuestas varía entre estudios: usar TF es ideal dado que la duración de los eslabones es constante sin importar la conducta, pero algunos estudios usan IF, lo que extiende la demora. Otros estudios han utilizado VI, lo que complica el modelamiento dado que se ha encontrado que los organismos no tratan a los programas VI y FI de la misma forma.

2.2. Duration of the S^- terminal link

Manipular la duración de los eslabones no reforzados independientemente parece tener poco efecto en la elección subóptima y en la elección en general: valores extremos no parecen afectar las preferencias.

2.3. Initial-Link duration and schedule

Suele haber mayor elección subóptima con programas largos de VI que con FR 1.

2.4. Duration of the intertrial interval

El ITI parece tener un efecto casi nulo en la elección: Spetch *et al.* (1990) no encontraron efectos con intervalos que variaron entre 0 y 40s.

3. Probability of food on each alternative

Disminuir la probabilidad de comida debajo de 1.0 en la alternativa óptima lleva a mayor elección subóptima, suponiendo que sea no señalada. Disminuir la probabilidad de la alternativa subóptima afecta poco la elección, mientras se mantenga por encima de .1.

4. Probabilities of differing amounts

Procedimiento de magnitudes: palomas prefieren la alternativa subóptima todavía.

5. Exposure to the contingencies

Algunos estudios usan criterios de estabilidad en lugar de números fijos de sesiones, lo que lleva a cantidades distintas entre sujetos. Estas diferencias en el entrenamiento pueden llevar a diferencias en elección, dado que se ha encontrado que la suboptimalidad incrementa con el entrenamiento.

Existen diferencias entre estudios en la cantidad de ensayos forzados. Estudios sin ensayos forzados tienen proporciones de elección relativamente bajas, y viceversa. Belke y Spetch (1994) usaron un procedimiento en el que se repetía todo ensayo que no terminaba en comida, forzando a las palomas a permanecer en la alternativa subóptima hasta obtener comida. Esto incrementó la preferencia por la alternativa subóptima.

6. Other variables

No parece importar que las alternativas no señaladas tengan un solo estímulo o dos.

Mayor privación de alimento y entornos pobres llevan a mayor elección subóptima.

La modalidad de la respuesta parece ser importante: se ha encontrado que las palomas tienen preferencia por la alternativa óptima cuando el operando de respuesta es un pedal.

7. Nonavian species

En ratas se han encontrado resultados contradictorios. Chow sugiere que la luz y tono usados como estímulos no adquieren saliencia incentiva dado que evocan conducta de goal tracking más que sign tracking. Zentall sugirió que la elección subóptima en ratas puede depender de su localización en su sistema conductual y si el estímulo evoca conducta de búsqueda general o focal. Sugiere que los estímulos visuales evocan conducta focal en palomas, pero no en ratas. Cunningham y Shahan sugieren que factores temporales importan, y encontraron más suboptimalidad con TL largos. Martínez sugirió que la inhibición condicionada es importante, y encontró optimalidad cuando una palanca fue usada como S^-. Alba encontró suboptimalidad con TL largos solo cuando los estímulos eran compuestos de tono más luz.

Las ratas son menos propensas que las palomas a la elección subóptima, pero su suboptimalidad parece incrementar con el TL.

7.1. Humans

Lalli (2000) evaluaron a niños con retraso en el desarrollo y encontraron resultados similares a los de palomas. Molet encontró una preferencia modesta por la alternativa

subóptima en jugadores. McDevitt encontró preferencia óptima en un procedimiento en humanos con la versión de magnitudes, igual que Stagner (2020).

Hay evidencia de que los humanos eligen la información anticipada aun cuando no tiene efecto en la recompensa final, y la preferencia por la información incrementa con la demora entre la elección y la recompensa. Algunos experimentos en humanos describen las contingencias de manera explícita, pero hay evidencia que indica que la elección es distinta cuando la contingencias se describan comparada con cuando se aprenden por experiencia. Aun así, los efectos similares de la duración de la demora y el énfasis en las buenas noticias indican aspectos en común entre la elección subóptima animal y la búsqueda de información en humanos.

8. Summary and implications

En palomas y estorninos los estímulos predictivos refuerzan la elección cuando la entrega de comida es probabilística y demorada. Las palomas eligen las alternativas señaladas incluso si su probabilidad de reforzamiento es sustancialmente menor que la probabilidad de la alternativa no señalada. El nivel de preferencia depende en gran medida de aspectos temporales: eslabones terminales largos y eslabones iniciales cortos promueven la elección subóptima. Ni los ITI ni la duración del S^- afecta a la elección.

Las ratas tienen menos probabilidad que las palomas de elegir la alternativa subóptima, pero su elección incrementa en función del TL.

Una hipótesis es que la elección por alternativas señaladas es determinada por la reducción de la incertidumbre. Evidencia en contra se encuentra en el hallazgo de que un S^- no sostiene las respuestas de observación a pesar de reducir la incertidumbre. Un experimento de Sears en el que una alternativa de buenas noticias fue preferida sobre una de malas noticias es también evidencia en contra.

Otra hipótesis es que solo el valor predictivo de S^+ determina la elección, y ésta no es afectada por su frecuencia o la frecuencia del reforzamiento primario. Pero en condiciones de 50 % vs 100 % se prefiere la alternativa señalada de 50 % a pesar de que los estímulos no discriminativos sean igualmente predictivos de la demora a la comida.

La diferencia en probabilidades de reforzamiento por sí misma tampoco explica la elección, es decir, el incremento en probabilidad señalado por S^+ no parece ser el determinante de la preferencia. El papel que tienen lo IL y TL indica que factores temporales también están implicados.

Las hipótesis de un solo mecanismo no parecen explicar la elección subóptima, por lo que muchas teorías apelan a más de un proceso. El modelo SiGN asume que la elección es determinada por la interacción entre el reforzamiento primario y el condicionado provisto por la reducción de la demora.

9. The SiGN model

9.1. Assumptions and formalization

Se asume que la elección es determinada por la competencia entre reforzamiento primario y condicionado con base en la teoría de reducción de la demora. La hipótesis se formaliza así:

$$\begin{array}{rcll} \frac{R_a}{R_a+R_b} & = & \frac{r_a(T-t_a)}{r_a(T-t_a)+r_b(T-t_b)} & \text{when } t_a < T, t_b < T \\ & = & 1 & \text{when } t_a < T, t_b > T \\ & = & 0 & \text{when } t_a > T, t_b < T \end{array}$$

donde R son las respuestas por alternativa, r son las tasas de reforzamiento primario, y $(T-t)$ son las reducciones en la demora, donde T es el tiempo medio global programado hasta el reforzamiento primario desde el inicio del eslabón inicial (sin contar ITI), y t son los tiempos al reforzamiento primario desde el inicio de los eslabones terminales respectivos. En este y otros modelos se hacen los cálculos con los valores programados y no los obtenidos.

Esta ecuación solo puede ajustarse a situaciones donde todo eslabón inicial es seguido de un solo eslabón terminal que siempre lleva a reforzamiento, pero el principio es generalizable. Una aproximación de Dunn y Spetch lleva esa lógica a situaciones probabilísticas describiendo cómo la probabilidad afecta a la demora. El tiempo al reforzamiento primario desde el eslabón terminal no señalado está dado por la expresión $p_a\dot{t}_a + (1-p_a)(t_a+T)$, donde p_a es la probabilidad de comida al final del eslabón, es decir, con probabilidad $(1-p_a)$ el tiempo incrementa en T dado que la recompensa se omite. Así, la reducción promedio en demora señalada por los estímulos de una alternativa probabilística no señalada está dada por

$$D_{avg} = T - (p_a \cdot t_a + (1-p_a)(t_a+T)),$$

lo que se simplifica a

$$D_{avg} = T \cdot p_a - t_a.$$

Cuando una alternativa es no señalada, p y t son los valores promedio de la alternativa. Cuando los resultados son señalados, el estímulo con mayor probabilidad de reforzamiento relativa a su demora (i.e., $\frac{p}{t}$) provee una reducción de demora *bonus* (D_{bonus}) relativa al tiempo promedio señalada por el inicio del eslabón terminal. Este bonus son las "buenas noticias". Sin embargo, las "buenas noticias" se definen en términos de los parámetros experimentales y no se refieren a la "información". Es decir, se trata de una descripción funcional.

Este modelo es especial porque incorpora la idea de que la reducción opera entre e intra alternativas (en el caso de eslabones señalados). Esas son las demoras promedio y *bonus* que llevan a la reducción *total* dentro de una alternativa. La reducción *bonus* del S^+ es la reducción promedio de un eslabón en esa alternativa menos la demora señalada por el S^+:

$$D_{\text{bonus}} = T \cdot (p_{S^+} - p) - t_{S^+} + t.$$

El modelo asume que el tiempo pasado en presencia de S^- no tiene ningún efecto más allá del tiempo que lleva a los organismos percibir el estímulo (se asume alrededor de 1 s). Esto afecta solamente la tasa de reforzamiento primario r y el tiempo promedio T en la alternativa señalada.

10. Balance of conditioned and primary reinforcement

En elección subóptima los reforzamientos primario y condicionado compiten en función de la duración de los eslabones terminales. Incrementar la duración del TL reduce la eficacia del reforzamiento primario.

La forma en que los TL afectan a la reducción promedio en la demora es capturada por

$$\beta = \log_{10}\left(1 + \frac{t_{S^+}}{i}\right).$$

β describe cómo el reforzamiento condicionado de una alternativa incrementa o disminuye con base en las duración de IL y TL en una alternativa señalada. TL largos incrementan el efecto de reforzamiento condicionado de la reducción *bonus*, mientras que IL largos lo disminuyen. Dado que no hay reducción bonus en alternativas no señaladas, β no afecta las predicciones de la teoría de reducción de la demora, y mantiene la propiedad de ser temporalmente relativa: el contexto temporal relativo determina la preferencia, por lo que no importa la escala temporal específica.

11. Predicting choice proportions

El modelo SiGN puede declararse formalmente así:

$$\begin{aligned} \frac{R_a}{R_a+R_b} &= \frac{r_a\delta_a}{r_a\delta_a+r_b\delta_b} && \text{when } \delta_a > 0, \delta_b > 0 \\ &= 1 && \text{when } \delta_a > 0, \delta_b < 0 \\ &= 0 && \text{when } \delta_a < 0, \delta_b > 0, \end{aligned}$$

donde δ representa el reforzamiento condicionado de una alternativa determinado por la suma de la reducción promedio de demora, D_{avg}, y la reducción bonus, D_{bonus}, ajustada por β.

$$\delta = D_{\text{avg}} + \beta \cdot D_{\text{bonus}}.$$

En esta forma el modelo puede cambiarse a una razón con parámetros de sesgo y sensibilidad.

De forma general, la tasa de reforzamiento primario en una sola alternativa es $r = \frac{1}{IL+TL}$, donde IL y TL representan al tiempo promedio pasado en eslabones iniciales y terminales por reforzador primario, y el 1 representa a un solo reforzador. Por ejemplo, en la alternativa discriminativa de elección subóptima se asume que los animales responden en el IL en 1s. Dado que la probabilidad de reforzamiento es de .2, las palomas deben pasar en promedio 5 veces por el IL para obtener un reforzador, lo que es un total de 5s esperando por comida en el eslabón inicial.

El tiempo de espera en el eslabón terminal se calcula de forma similar, pero integrando las dos posibilidades que el organismo se puede encontrar: para $a1$, que ocurre con probabilidad de .2, hay un eslabón de 10s. Para $a2$, que ocurre con probabilidad .8, se asume un eslabón de 1s, dado que se supone que las palomas no toman en cuenta ese tiempo. En promedio, el tiempo pasado en un eslabón terminal es de $0{,}2 \times 10 + 0{,}8 = 2{,}8s$. El tiempo promedio esperando por comida en la fase de eslabón terminal es de $2{,}8 \times (1/0{,}2) = 14s$, donde 0.2 es la probabilidad de reforzamiento de la alternativa. La tasa de reforzamiento de la alternativa A es el inverso del tiempo total pasado en los eslabones iniciales y terminales de la alternativa: $r_a = \frac{1}{5+14} = 0{,}053$. Aplicar la misma lógica a la alternativa B, el tiempo promedio pasado en eslabones iniciales por entrega de comida es de 2s, y el tiempo promedio pasado en eslabones terminales es de 20s. Por lo tanto, $r_b = \frac{1}{2+20} = 0{,}045$.

Con r_a y r_b fijos, el tiempo promedio programado hasta el reforzamiento primario desde el inicio de un eslabón inicial es:

$$T = \mathbf{P}_a \frac{1}{r_a} + \mathbf{P}_b \frac{1}{r_b},$$

donde $\mathbf{P}_a$ y $\mathbf{P}_b$ son las proporciones programadas de entrega de comida en cada alternativa con relación a la otra.

En el ejemplo de .2 vs .5: $\mathbf{P}_a = \frac{0{,}2}{0{,}2+0{,}5} = 0{,}286$, donde .2 y .5 son las probabilidades de reforzamiento en las alternativas A y B. $\mathbf{P}_b = 1 - \mathbf{P}_a$. El cálculo de T resulta entonces en 21.143s.