

La elección en humanos como una función del valor subjetivo y distancia temporal de las experiencias previas

Luis J. López-Romero*, Angélica Alvarado, Rosalva Cabrera, David Luna, Javier Vila

F. E. S. Iztacala Universidad Nacional Autónoma de México, México

ABSTRACT

Choice in humans as a function of subjective value and temporal distance of previous experiences. Choice behavior has been studied to determine how past experiences are integrated to affect elections between patches. This idea suggests an integration of subjective value and temporal distance in a dynamic average to predict choice behavior. A choice task was presented on a computer screen, which showed three options that could give or not an appetitive consequence. Participants were divided into 4 groups depending on either reward magnitude or training-test interval. Results showed: 1) preference for the most recently reinforced option (B) when test was immediate, 2) preference for the option with the highest reward magnitude (A) at 24h (spontaneous recovery) in the group $A > B$ 24h, 3) indifference in $A = B$ 24h groups, consistent with idea that subjective values and recency of past experiences change as time passes. These results are consistent with the predictions of Temporal Weighting Rule (TWR) and assume integration of information based on dynamic average of past experiences in human choice tasks.

Key words: choice behavior, temporal weighting rule, integration of information, humans.

RESUMEN

La conducta de elección puede estudiarse como una integración de experiencias pasadas en la elección entre varias opciones. Esta idea integra el valor subjetivo y la distancia temporal relativa de las experiencias anteriores en un promedio dinámico para predecir la conducta de elección. Se usó una tarea de elección presentada en una pantalla de ordenador en donde se mostraban tres opciones de respuesta que podían otorgar o no una consecuencia apetitiva. Los participantes se asignaron a 4 grupos que se diferenciaban en la magnitud de la recompensa o en el valor del intervalo entrenamiento-prueba. Los resultados mostraron: 1) preferencia por la opción más recientemente reforzada (B) en la prueba inmediata, 2) preferencia por la opción con mayor magnitud de recompensa (A) a las 24h (recuperación espontánea), grupo $A > B$ 24h y 3) indiferencia en la preferencia, grupos $A = B$ 24h, lo que sugiere que los valores subjetivos y de recencia de las experiencias pasadas cambian con el tiempo. Estos resultados son coherentes con las predicciones de la Regla de Peso Temporal (RPT) y suponen que en una tarea de elección con humanos las experiencias pasadas se integran en base a un promedio dinámico.

Palabras clave: conducta de elección, regla de peso temporal, integración de información, humanos.

* Correspondencia a: Dr. Javier Vila Carranza, División de Investigación y Posgrado, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México. Tlalnepantla, Edo Mex. 54096, México. Email: javila@campus.iztacala.unam.mx. Este estudio fue financiado por el proyecto DGAPA, IN304513. de la UNAM.

El estudio de la conducta de elección se refiere a los factores que hacen que los organismos realicen una u otra conducta ante varias alternativas. En algunos estudios con humanos (Hastie y Dawes, 2001) se hace referencia a una “deliberación racional” para describir situaciones donde se reflexiona acerca de diferentes alternativas, como cuando tratamos de decidir entre dos alimentos o elegir rutas alternas para evitar el tráfico. Sin embargo, tanto en los estudios con animales como con humanos es difícil observar esta racionalización, lo que hace complicado el estudio de la elección; aún así existen investigaciones que han tratado de estudiar si los mecanismos psicológicos involucrados son similares entre especies.

En la literatura de condicionamiento operante el estudio de la elección se ha llevado a cabo a partir del estudio de programas concurrentes en los que dos o más programas de reforzamiento están disponibles simultáneamente cada uno con un manipulando diferente, y el sujeto o participante es libre de distribuir sus respuestas entre las alternativas disponibles (Herrnstein, 1961). A partir de estos programas se han desarrollado diferentes procedimientos experimentales y modelos teóricos para explicar la conducta de elección en diferentes situaciones experimentales. Por ejemplo, la ley de igualdad (Davison y McCarthy, 1988; Williams, 1988), el modelo de elección contextual (Grace, 1994; 1996), o la teoría de la reducción de la demora (Fantino, 1969). Estos modelos hacen diferentes suposiciones sobre los mecanismos de la conducta de elección y cada uno ha proporcionado resultados significativos con un gran número de experimentos en programas concurrentes (Mazur, 2001).

La ley de igualdad de Herrnstein (1961) plantea una relación funcional entre la conducta medida como tasa de respuestas y algún parámetro de reforzamiento (frecuencia relativa, magnitud, demora) en una expresión cuantitativa. Planteada a partir de situaciones en las que más de una respuesta es reforzada, y en las que la tasa de respuesta relativa para una alternativa de respuesta en particular es igual a la tasa de reforzamiento relativo para esa respuesta. Esta relación de igualdad es considerada actualmente como una regla empírica de la elección bajo programas de reforzamiento concurrentes IV-IV.

Una de las implicaciones de esta ley es que la respuesta no sólo depende de su tasa de reforzamiento específica, sino que interactúa con las tasas de reforzamiento de otras conductas. Así, los mecanismos que explican la conducta de elección desde la perspectiva de la ley de igualdad consideran ya sea el promedio de las tasas de reforzamiento y de respuesta a lo largo de la toda la sesión de entrenamiento (teorías molares; Rachlin, Green, Kagel y Battalio, 1976), o en el resultado neto de cada elección individual (teorías moleculares; Shimp, 1969; Williams, 1992).

En su ambiente natural los animales tienden a visitar diferentes lugares para obtener alimento que puede estar disponible en diferentes cantidades, lugares y momentos, que además se puede deteriorar o cambiar por las condiciones ambientales variables. Por lo tanto deben utilizar una estrategia de búsqueda que les permita elegir de manera óptima aquellas parcelas que les lleven a obtener los mayores beneficios. Así, algunos estudios recientes sobre las estrategias de forrajeo en animales que incluyen situaciones de elección entre diferentes opciones de respuesta han sugerido que el recuerdo de las experiencias pasadas puede combinarse e influir en la conducta presente. De acuerdo

a esta idea los cambios en la elección entre diferentes opciones de respuesta pueden explicarse si consideramos que se aprende la información de cada experiencia por separado, y luego se realiza un promedio de su valor subjetivo al inicio de una experiencia nueva, de tal forma que la elección se basa en la integración de toda las experiencias aprendidas previamente (Mazur, 1995; Devenport y Devenport, 1994).

Mazur (1996) ha estudiado esta posible integración de información de experiencias previamente reforzadas durante varias sesiones en una última elección entre varias alternativas. En su estudio, registró las respuestas de palomas en un programa concurrente de reforzamiento en el que cambiaba el porcentaje de reforzamiento relativo a cada uno de los componentes del programa de manera sucesiva durante varias sesiones. En una primera condición, una de las opciones era “más rica” en promedio que la otra, y en una segunda condición ambas opciones eran iguales en promedio durante todas las sesiones restantes. Los resultados mostraron una preferencia en la última sesión por aquella opción de respuesta que proporcionaba mayor reforzamiento, y mostraron indiferencia en la elección cuando ambas opciones de respuesta proporcionaban la misma cantidad de reforzamiento. Sin embargo, esta preferencia cambiaba después de un intervalo de 24h en la condición en la que una de las opciones era “más rica” en promedio que la otra. Pero en la condición en la que ambas opciones de respuesta tenían el mismo reforzamiento durante todas las sesiones, en la prueba de elección después de 24h se observó indiferencia entre las opciones de respuesta. Lo que sugiere que la elección de las palomas después de 24h se realizaba en base a la opción que en promedio proporcionaba mas reforzadores, mientras que la elección inmediata siempre ocurrió en base a la opción que recientemente había recibido más reforzamiento. Resultados similares han sido observados en humanos en un experimento similar (López-Romero, Alvarado, Tamayo y Vila, 2011). El cambio en la elección por la opción más rica en promedio luego de un intervalo de retención, ha sido interpretado como una integración de la información de lo ocurrido en todas las sesiones (Devenport, 1998; Mazur, 1996).

Esta integración de información esta basada en dos parámetros básicos de las experiencias de aprendizaje: su distancia temporal relativa (recencia) y su valor subjetivo (Devenport y Devenport, 1994). La Regla del Peso Temporal (RPT) expresa formalmente una relación inversa entre estos factores (cuando el peso de la recencia es mayor el peso del valor subjetivo es menor y viceversa) en la siguiente ecuación:

Donde, vw indica el valor de estimación temporal de una experiencia, Q el valor

$$(1) \quad vw = \frac{\sum_{n=1}^{n=j} (Q * 1/T)}{\sum_{n=1}^{n=j}}$$

subjetivo estimado y $1/T$ el valor de recencia. Así, cuando se requiere de una respuesta de elección entre dos alternativas, se debe calcular el valor de estimación temporal de una segunda experiencia (vwB) además de tomar en cuenta la estimación temporal total de ambas experiencias, lo que se expresa en la ecuación 2. En donde la elección de la opción B es la proporción de vwB en relación a vwA : (2) $vwB = vwB / vwA + vwB$

Las predicciones básicas de la RPT (Devenport, Hill, Wilson y Odgen, 1997) suponen que durante la última elección entre dos experiencias se promedian los pesos de

la distancia relativa y del valor subjetivo de cada alternativa. Inicialmente las experiencias recientes obtienen el mayor peso, pero al pasar tiempo y perderse la recencia, será el valor subjetivo de cada alternativa lo que tendrá entonces un mayor peso y determine entonces la elección observada. La RPT supone así la integración de las experiencias pasadas en base a estas dos variables; su recencia o distancia temporal relativa ($1/T$) y su valor subjetivo (Q). Que en la ecuación 1 es representada como ($Q * 1/T$), el cual cambia con el paso del tiempo haciendo que Q aumente al disminuir $1/T$, lo que implica que las experiencias tienden a igualarse al perder su recencia, por lo que se observará la elección de aquella experiencia que haya tenido una mayor valor subjetivo (Q).

Una primera predicción básica derivada de estas ideas es que al aumentar el intervalo de retención después del entrenamiento sucesivo de dos experiencias A y B la elección de cada una de ellas dependerá de la recencia de cada una y de su valor subjetivo. Así, cuando el valor subjetivo de A es mayor que B; en una prueba inmediata se elegirá la opción reforzada recientemente, sin embargo al aumentar el intervalo de retención entrenamiento-prueba se observará un efecto recencia-primacía en el que la preferencia por la opción A incrementa proporcionalmente al aumentar el intervalo de retención en una prueba demorada, debido a que el valor subjetivo de A es mayor en promedio que el de B (Vila, López-Romero y Alvarado, 2010). Una segunda predicción es que cuando ambas experiencias tienen el mismo valor subjetivo ($A=B$), en una prueba inmediata se preferirá la opción B por ser la más reciente, sin embargo al aumentar el intervalo de retención se observará indiferencia entre ambas opciones, debido a que el valor subjetivo de ambas opciones es el mismo al perderse la recencia de B (López-Romero, Garcia-Barraza y Vila, 2010).

La presente investigación evaluó la posibilidad de observar estas predicciones de la RPT en una tarea de elección con humanos empleando una tarea de condicionamiento instrumental similar a la empleada por Vila *et al.* (2010). Así, para demostrar el supuesto de la RPT en humanos de que las experiencias previas son integradas en un promedio dinámico basado en su distancia temporal relativa y su valor subjetivo, que determina la elección entre varias opciones de respuesta, se estudió una situación experimental con tres opciones de respuesta. De este modo, en un experimento con humanos se varió tanto la distancia temporal relativa en dos pruebas (0 y 24h) como el valor subjetivo de cada experiencia (mayor o menor reforzador). Para ello cuatro grupos recibieron un entrenamiento en dos fases sucesivas en donde tres opciones de respuesta cambiaban de valor en cada fase y finalmente se realizó un prueba inmediata (0h) o distante (24h) de elección entre las opciones de respuesta.

MÉTODO

Participantes

Participaron voluntariamente 48 estudiantes universitarios de entre 19 y 21 años de edad (32 mujeres y 16 hombres), alumnos del Instituto Politécnico Nacional y de la Facultad de Estudios Superiores Iztacala de la UNAM de psicología sin experiencia previa en este tipo de tareas y sin conocimientos de elección. Los participantes fueron

informados acerca de las condiciones éticas del estudio y por su participación recibieron créditos en una de las asignaturas que cursaban.

Aparatos y situación experimental

El experimento tuvo lugar en un cubículo de la FES Iztacala de 1,5x2m aproximadamente, que contaba con luz artificial, el cual se encontraba en el interior de la institución, aislado de ruidos externos y en cuyo interior se encontraba una silla y una mesa; así como en un salón de clases del Instituto Politécnico Nacional de 5x8m aproximadamente que contaba con sillas-bancos, una mesa e iluminación artificial. Un ordenador portátil *Gateway* que tenía instalado el programa *Super Lab Pro for Windows* v 4.0.2 (Cedrus Co.) fue utilizado para programar la tarea, presentar los estímulos, resolver la tarea y registrar las respuestas.

Tarea experimental

La tarea fue similar a la empleada por Vila *et al.* (2010), y consistió en una situación donde los participantes tenían que escoger entre tres opciones de respuesta que se presentaban de manera alterna y que se diferenciaban por la proporción de ensayos reforzados-no reforzados. La pantalla de un ordenador portátil simulaba ser una superficie de una mesa de juego que mostraba tres máquinas dispensadoras de cartas, una a la izquierda (azul), otra al centro (naranja) y la tercera a la derecha (amarilla) que daban una carta al pulsar el cursor del ratón sobre el botón "Pedir carta". Todas las máquinas aparecían simultáneamente en cada ensayo pero sólo una se encontraba activa y las otras dos permanecían "bloqueadas" por un círculo rojo con una diagonal que forzaba la respuesta a la máquina que se encontraba disponible en cada ensayo (ver figura 1). La posición de las máquinas que otorgaban puntos en alguna de las fases de entrenamiento fue contrabalanceada. La mitad de cada grupo tenía la máquina A a la izquierda y la B a la derecha mientras que en la otra mitad la posición se invirtió. La tarea de los participantes consistió en activar mediante *clicks* del ratón las máquinas hasta obtener una "carta ganadora" que otorgaba puntos bajo un programa de intervalo variable 5s (ver figura 1). Finalmente, se presentó una fase de prueba donde las tres máquinas fueron presentadas simultáneamente durante 30s sin reforzamiento, divididos en dos bloques de 15s, en los que la posición de las máquinas que otorgaban puntos cambiaba de lugar para evitar sesgos de preferencia de lugar. Los participantes tenían que hacer una elección entre las opciones de respuesta A, B y C respondiendo con *clicks* sobre cada máquina.

Procedimiento

La asignación de los participantes a cuatro grupos fue realizada de manera aleatoria. Los participantes tomaron asiento frente al ordenador y, a continuación, en una primera pantalla aparecían las instrucciones que debían seguir para realizar la tarea. Una vez que el experimentador comprobó que no existían dudas la sesión experimental

RESULTADOS

La medida utilizada para evaluar la preferencia de los participantes fue el porcentaje de elección de cada localización durante el primer bloque de 15s de la prueba. De esta manera se analizó si la preferencia de los participantes fue un efecto de recencia (elegir la alternativa más recientemente reforzada, es decir, B) o un efecto de primacía (elección de la alternativa de mayor magnitud de reforzamiento en promedio, es decir, A), o en su caso de indiferencia (elegir las opciones por igual).

Para analizar los resultados obtenidos se midió la preferencia entre las máquinas como el porcentaje de respuestas presentadas a la máquina A durante la prueba. Debido a que los participantes en ningún caso eligieron la opción C esta opción de respuesta no se tomó en cuenta para los análisis.

Se realizó un ANOVA 2 (Grupo) X 2 (Intervalo de retención) el cual reveló efectos principales para el factor Intervalo, $F(1,36) = 7.93, p < .05$, y no se observó ningún otro efecto significativo ($p > .05$). Los resultados muestran una preferencia por la máquina B en la prueba a las 0 horas en los grupos A= B 0h y A> B 0h y una preferencia por la máquina A después de 24 horas en el grupo A> B 24h, así como indiferencia en la elección del grupo A= B 24h.

Los resultados obtenidos durante las pruebas a 0 y 24 horas se presentan en la figura 2 y muestran: a) un efecto de recencia en los grupos A= B 0h y A> B 0h, mostrado una preferencia por la alternativa B, es decir, aquella que fue reforzada recientemente; b) indiferencia en la elección de las máquinas A y B para el grupo A= B 24h; c) Preferencia por la alternativa A en el grupo A> B 24h lo que sugiere tal y como lo predice la RPT, que la elección de los participantes se basó en la integración de información de la distancia temporal relativa y los valores subjetivos de cada opción de respuesta en todas las fases de entrenamiento.

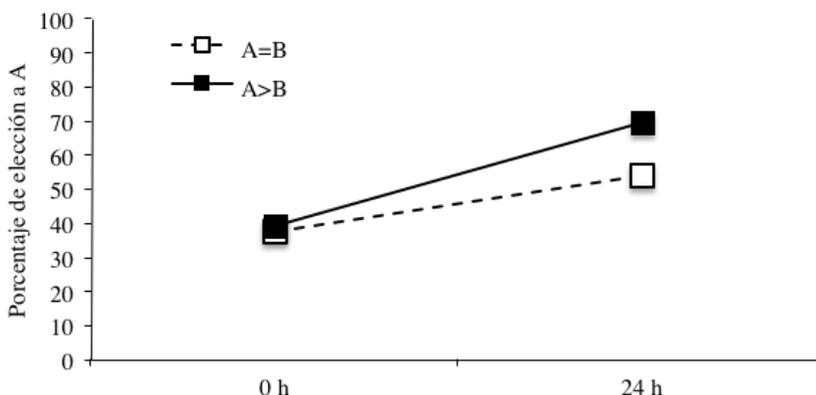


Figura 2. Porcentaje de respuestas a la alternativa A (alternativa reforzada mas distante) durante los diferentes intervalos entrenamiento-prueba (0h vs 24h). Los resultados muestran una preferencia por B (alternativa recientemente reforzada) en la prueba a las 0h, pero indiferencia a las 24h cuando las dos alternativas tienen el mismo valor subjetivo (Grupo A= B 24h) y una preferencia por A cuando esta tiene un valor subjetivo mayor (Grupo A> B 24h).

DISCUSIÓN

Tras un entrenamiento donde los participantes eligieron entre tres opciones (A, B y C) que se diferenciaban en la recencia al momento de la prueba (0 vs. 24h) o en su valor subjetivo ($A > B$, $A = B$ y C nada) se observó que en una prueba inmediata, eligen la alternativa reforzada recientemente, pero con el paso del tiempo el valor de la recencia se pierde en favor del valor subjetivo lo que produce una preferencia por la alternativa que en promedio otorgó mayor reforzamiento.

Los resultados obtenidos apoyan la idea de que en la conducta de elección, en una prueba inmediata, la distancia temporal es el parámetro de mayor peso, por lo que la elección se basa en lo que se aprende en la última fase, mientras que en una prueba demorada, la distancia temporal para las diferentes experiencias de aprendizaje, se iguala dejando al valor subjetivo como el parámetro de mayor peso para determinar la elección (Devenport y Devenport, 1994; Devenport *et al.* 1997). Estos resultados son coherentes con los resultados de estudios anteriores con humanos que sugieren también, que la elección se realiza con base en la distancia temporal relativa y al valor subjetivo de cada experiencia de aprendizaje. Vila, López y Alvarado (2010) observaron en una tarea similar que al aumentar en varios valores el intervalo de prueba posterior al entrenamiento se observó un cambio recencia primacía en el que la preferencia por la opción de mayor valor promedio es directamente proporcional al valor del intervalo. Mientras que López Romero, García Barraza y Vila (2010) observaron indiferencia entre las alternativas de respuesta al aumentar el intervalo de retención cuando ambas opciones tenían el mismo valor subjetivo, tal y como predice la RPT (Devenport y Devenport, 1994; Devenport *et al.* 1997).

De acuerdo con la ecuación (1), los presentes resultados muestran que con dos valores de distancia temporal relativa o recencia ($1/T$) en una prueba inmediata el peso más grande lo tiene la opción en la que la recencia es mayor, por lo que ambos grupos ($A = B$ 0h y $A > B$ 0h) en la prueba inmediata eligieron la opción B independientemente del mayor valor subjetivo (Q) de A, por lo que se predice una elección similar para B en ambos grupos, que fue lo observado en el presente experimento. Por otro lado, de acuerdo con la expresión ($Q * 1/T$) el valor subjetivo aumenta al disminuir la recencia, por lo que al pasar el tiempo el valor subjetivo (Q) será mayor y determinará la elección de la opción donde (Q) es mayor, tal y como se observó en el grupo $A > B$ 24h. De acuerdo con esto, al reducir la recencia de B ($1/T$) sólo queda el parámetro del valor subjetivo (Q), pero en el caso del grupo $A = B$ 24h las experiencias A y B son iguales en el valor de (Q) por lo que se observó indiferencia entre ellas. Finalmente tomando en cuenta lo anterior, la elección de los participantes es acorde con la ecuación (2) al considerar la preferencia por B (vwB) de acuerdo al valor subjetivo y distancia temporal relativa de las alternativas reforzadas A y B ($vwB/vwA + vwB$). Así, los resultados de la presente investigación al ser coherentes con las predicciones de la RPT, pueden ser interpretados como la integración de información de los valores subjetivos y de su recencia de las opciones de la conducta de elección, y sugieren a la RPT como una propuesta teórica útil para el estudio de la conducta de elección que permite integrar datos de animales y humanos a partir de principios similares.

REFERENCIAS

- Davison M y McCarthy D (1988). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.
- Devenport LD (1998). Spontaneous recovery without interference: Why remembering is adaptive. *Animal Learning and Behavior*, 26, 172-181.
- Devenport LD y Devenport JA (1994). Time-dependent averaging of foraging information in least chipmunks and golden-mantled ground squirrels. *Animal Behavior*, 47, 787-802.
- Devenport L, Hill T, Wilson M y Ogden E (1997). Tracking and averaging in variable environments: A transition rule. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 450-460.
- Fantino E (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior*, 12, 723-730.
- Grace R (1994). A contextual model of concurrent-chains choice. *Journal of the Experimental Analysis of Behavior*, 61, 113-129.
- Grace R (1996). Choice between fixed and variable delays to reinforcement in the adjusting-delay procedure and concurrent chains. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 362-383.
- Hastie R y Dawnes RM (2001). *Rational choice in an uncertain world*. Sage Publications. London, UK.
- Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of Experimental Analysis of Behavior*, 4, 267-272.
- López Romero LJ, Alvarado A, Tamayo C y Vila J (2011). La integración de experiencias pasadas y recientes determina la elección en humanos, *Revista Mexicana de Investigación en Psicología*, 3, 113-120.
- López Romero LJ, García Barraza R y Vila J (2010). Spontaneous recovery in human instrumental learning: Integration of information and recency to primacy shift. *Behavioral Processes*, 84, 617-621.
- Mazur JE (1995). Development of preference and spontaneous recovery in choice behavior with concurrent variable-interval schedules. *Animal Learning & Behavior*, 23, 93-103.
- Mazur JE (1996). Past experience, recency, and spontaneous recovery in choice behavior. *Animal Learning & Behavior*, 24, 1-10.
- Mazur JE (2001). Hyperbolic Value Addition and General Models of Animal Choice. *Psychological Review*, 108, 106-112.
- Rachlin H, Green L, Kagel JH y Battalio RC (1976). Economic demand theory and psychological studies of choice. In G Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 10, pp. 129-154). New York: Academic Press.
- Shimp CP (1969). Optimum behavior in free-operant experiments. *Psychological Review*, 76, 97-112.
- Vila J, Lopez, LJ y Alvarado A (2010). Spontaneous recovery as a dynamic average of previous experiences in human instrumental conditioning. *International Journal of Psychology and Psychological Therapy*, 10, 403-413.
- Williams BA (1988). Reinforcement, choice, and response strength. In RC Atkinson, RJ Herrnstein, G Lindzey, y RD Luce (Eds.), *Stevens' Handbook of Experimental Psychology: Vol. 2. Learning and cognition* (2nd ed., pp. 167-244). New York: Wiley.
- Williams BA (1992). Dissociation of theories of choice via temporal spacing of choice opportunities. *Journal of Experimental Psychology: Animal Behavior Processes*, 18, 287-297.

Recibido, 9 Febrero, 2012

Aceptación final, 22 Febrero, 2013