

# Modelos bayesianos de inferencia psicológica: ¿cómo predecir acciones en situaciones de incertidumbre?\*

## Bayesian Models of Psychological Inference: How to Predict Actions under Uncertainty Situations?

Recibido: junio 3 de 2009 | Revisado: septiembre 20 de 2009 | Aceptado: octubre 1 de 2009

HERNANDO TABORDA\*\*

Universidad del Valle, Cali, Colombia

### RESUMEN

Esta investigación examinó si la capacidad para predecir acciones en circunstancias de incertidumbre puede ser comprendida como un proceso de inferencia bayesiana. Se llevaron a cabo dos experimentos, cada uno con un grupo diferente de estudiantes universitarios del ciclo básico. En ambos experimentos, se presentó un video en donde un agente entraba a un laberinto para consumir alimento. La tarea de los participantes fue predecir cuál de cuatro entradas utilizaría el agente para ingresar a la parte interna del laberinto. Los resultados mostraron una correlación alta, significativa a un  $p < 0.01$ , entre las predicciones del modelo bayesiano y las respuestas de los participantes. Este hallazgo revela que la capacidad de predecir acciones implica la estimación de la probabilidad posterior, de acuerdo con el modelo bayesiano propuesto.

### Palabras clave autor

Esquema teleológico, inferencia Bayesiana, modelos causales, predicción de acciones.

### Palabras clave descriptores

Experimentación con animales, teoría bayesiana de decisiones estadísticas, condicionamiento (psicología).

### ABSTRACT

This study investigated if capacity for predicting actions under uncertainty conditions can be understood like a Bayesian inferential processes. Two experiments were carried out, each one with a different group of basic cycle undergraduate students. In both experiments it was presented a film that showed an agent getting into maze for eating meal. Participants had to predict which of four entrances would use the agent for getting inside of the maze. Results showed a high correlation, significant to a  $p < 0.01$ , between predictions of a Bayesian inferential model and the answers of the participants. Findings reveal that capacity for predicting actions involve estimations of posterior probability in accordance with the Bayesian model proposed.

### Key words author

Bayesian Inference, Action's Prediction, Causal Models, Teleological Schema.

### Key words plus

Animal Experimentation, Bayesian Statistical Decision Theory, Conditioning (Psychology).

---

Para citar este artículo. Taborda, H. (2010). Modelos bayesianos de inferencia psicológica: ¿cómo predecir acciones en situaciones de incertidumbre? *Universitas Psychologica*, 9 (2), 495-507.

---

\* Artículo de investigación.

\*\* Centro de Investigaciones en Psicología, Cognición y Cultura, Universidad del Valle, Calle 13 # 100-00, Edificio 385, Oficina 4011, tel: 3212192. Cali, Valle, Colombia. Correo Electrónico: htabordao@yahoo.com.ar

## Introducción

La habilidad para predecir las acciones de otra persona es esencial, en múltiples escenarios de interacción social. Durante los últimos años, se han llevado a cabo varios estudios que intentan esclarecer el tipo de mecanismo cognitivo involucrado en este proceso inferencial (Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra & Bíró, 1995; Meltzoff, 1995).

Csibra y Gergely (2007) proponen un mecanismo que explica la realización de varios tipos de inferencia psicológica, incluida la predicción de acciones. Este mecanismo se fundamenta en la capacidad que tienen las personas, posiblemente desde el primer año de vida, para elaborar una *representación teleológica* de la acción. De acuerdo con esta propuesta, las personas construyen una representación “bien formada” de la acción mediante tres componentes: la intención, la meta y las restricciones físicas. Se considera “bien formada” sólo si los tres elementos se pueden coordinar mediante la aplicación del “principio de racionalidad” (Csibra, 2003; Csibra, Bíró, Koós & Gergely, 2003). Por lo tanto, la hipótesis según la cual las personas son agentes que actúan racionalmente, es fundamental para anticipar las acciones o predecir su meta buscada. Si el agente es racional, tratará siempre de realizar la acción más eficiente, dadas las restricciones propias de su ambiente.

Si bien el mecanismo de inferencia psicológica propuesto por Csibra y Gergely se encuentra bien fundamentado empíricamente, es importante notar la insuficiencia de su funcionamiento en múltiples circunstancias. Por ejemplo, Kording y Wolpert (2006) señalan que la inferencia de estados mentales depende en gran medida de la inferencia previa de la “función de utilidad” del agente observado, es decir, que para predecir la acción más eficiente es necesario inferir antes cuáles son los beneficios que un agente determinado privilegiaría y cuáles serían las consecuencias que trataría de evitar.

Otra limitación importante del modelo de Csibra es que la comprensión de las metas, las intenciones y las creencias, rara vez, se dan con un grado de certeza absoluto. Predecir una acción implica

conocer de antemano la meta y las restricciones, pero en algunas situaciones las personas sólo tienen un conocimiento aproximado o incierto de estos elementos. Siguiendo esta ruta, la pregunta primaria de la actual investigación es si las personas tienen la capacidad de predecir la siguiente acción de un agente, a partir de información probabilística; es decir, si pueden realizar inferencias psicológicas en situaciones de incertidumbre sobre el valor exacto de los elementos de la representación teleológica y, si es así, ¿cómo lo hacen?

Puede ser evidente que en múltiples circunstancias las personas deban razonar bajo condiciones de incertidumbre y que lo hagan de manera adaptativa, pero no es posible inferir de allí que las personas representen el valor probabilístico de cada parte del esquema teleológico y que operen con base en estas probabilidades. Sin embargo, investigaciones recientes sobre inferencia causal sugieren que la respuesta a este interrogante puede ser positiva.

Al igual que en el dominio psicológico, el pensamiento causal funciona en circunstancias de alta incertidumbre. La asociación entre causa y efecto no debe ser unívoca y determinista necesariamente, sino que puede funcionar con diferentes grados de probabilidad. La teoría de las Redes Bayesianas se creó en el seno de la estadística y la inteligencia artificial, para intentar modelar la manera como un sistema inteligente debería inferir relaciones causales, en este tipo de circunstancias. En años recientes, la investigación psicológica ha reunido evidencia que sugiere que las personas razonan de manera análoga a como lo hacen los modelos computacionales de las Redes Bayesianas (Glymour, 2003; Gopnik, Glymour, Sobel, Shulz, Kushnir & Danks, 2004; Gopnik & Schulz, 2004; Gopnik & Tenenbaum, 2007).

De acuerdo con el modelo de redes bayesianas la inferencia causal opera en dos estadios diferentes (Griffiths, Kemp & Tenenbaum, 2008; Griffiths & Tenenbaum, 2005). En primer lugar, la inferencia implica la *selección del modelo causal* correcto. Es decir, que los eventos observados pueden ser representados mediante un gráfico acíclico dirigido (DAG, por sus siglas en inglés), en donde cada variable es representada por un nodo y las relaciones

causales entre ellas, mediante flechas unidireccionales. Por lo tanto, la inferencia causal requiere determinar el número de variables del sistema y el tipo de relaciones condicionales entre ellas. La investigación computacional ha desarrollado una serie de algoritmos que tienen como propósito inferir el modelo causal correcto, a partir únicamente de la observación de las relaciones de probabilidad condicional entre los eventos (algunos algoritmos operan también con base en las intervenciones sobre el sistema).

El segundo estadio de la inferencia es la *estimación de parámetros* del modelo seleccionado. En este estadio se trata de determinar cuál es el peso probabilístico de cada uno de los nodos del modelo, condicional a la ocurrencia de aquellos nodos del sistema de los cuales depende (únicamente de ellos). La potencia de la aproximación bayesiana al problema de la estimación de parámetros, radica en que provee un algoritmo sencillo para actualizar, a través de observaciones sucesivas, el valor de los parámetros del sistema. Este algoritmo es conocido como el Teorema de Bayes. Mediante este teorema es posible determinar el valor de la probabilidad posterior de un evento o hipótesis particular, a partir de la información sobre su probabilidad previa y la información actualmente observada.

Las investigaciones sobre razonamiento causal, tanto en niños como en adultos, se han enfocado primordialmente en el problema de la selección de modelos (Gopnik & Schulz, 2004). Los resultados de estas investigaciones indican que, al menos desde los 2 años de edad, las personas son capaces de recuperar la estructura causal, a partir de la dependencia condicional observada entre los eventos. Otro grupo de investigaciones (Meder, Hagmayer & Waldmann, 2008; Waldmann & Hagmayer, 2005), muestran también que los adultos tienen la capacidad de predecir el peso de la probabilidad condicional de una variable de manera muy similar a como lo haría el teorema de Bayes, lo cual indica que pueden representar los parámetros del modelo y operar sobre ellos.

Recientemente, se han propuesto algunos desarrollos teóricos que permitirían extender el uso del modelo de redes bayesianas a otros dominios

de conocimiento fuera del causal. Tenenbaum, Griffiths y Kemp (2006), sostienen, por ejemplo, que la inferencia bayesiana estaría implicada en la generalización inductiva de conceptos, el aprendizaje de palabras y la inducción de propiedades biológicas.

En una serie de artículos, Baker, Tenenbaum y Saxe (2006, 2007) proponen que el modelo de “redes jerárquicas Bayesianas”, como ellos lo denominan, puede extenderse también al dominio del razonamiento psicológico y en específico al esquema teleológico de Csibra y Gergely. En los estudios reportados, los investigadores intentan determinar si la manera como las personas predicen la meta perseguida por un agente, se adecua a un modelo bayesiano. Con este propósito, crean un ambiente virtual en donde un círculo sigue una trayectoria errática hacia tres posibles metas. En algunos casos se muestra un pequeño muro entre la posición del “agente” y una de las metas. La tarea de los participantes, todos sujetos adultos, fue predecir en cinco puntos diferentes de la trayectoria, cuál es la meta que al parecer persigue el agente. Los resultados muestran que las personas pueden predecir la meta buscada mediante un proceso inferencial bayesiano. Resultados similares fueron obtenidos previamente por Rao, Shon y Meltzoff (2004) en el campo de la robótica.

## El presente estudio

La actual investigación pretende extender los resultados obtenidos hasta el momento, sobre la implementación de modelos computacionales bayesianos en distintos ámbitos de la cognición humana. En específico, pretende extender los estudios anteriores de tres maneras diferentes. En primer lugar, se quiere indagar en un proceso inferencial no abordado previamente en la literatura sobre modelos bayesianos, como es el proceso de predicción de acciones. De acuerdo con Csibra, éste es un proceso funcionalmente independiente de la predicción de metas. Por lo tanto, es importante preguntar si bajo condiciones de incertidumbre este proceso inferencial, se adecua en algún grado a un razonamiento bayesiano.

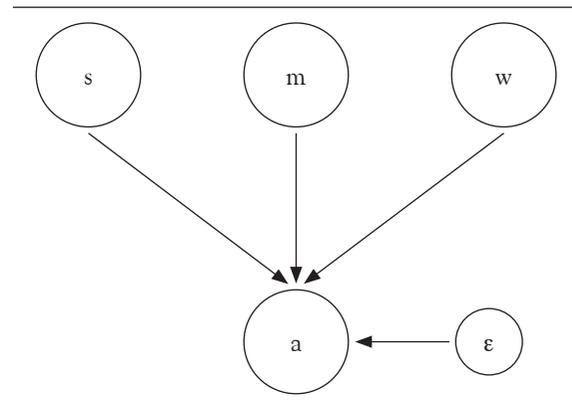
El segundo objetivo es determinar el grado de estabilidad y homogeneidad del uso de la evidencia de la probabilidad condicional. Los experimentos habitualmente diseñados estudian la parametrización del modelo causal a través de pocas predicciones. Como en el experimento de Baker et al. (2007) presentado atrás. Mediante un conjunto más amplio y complejo de predicciones esperamos observar con más detalle si las personas recuperan el conjunto total de la probabilidad conjunta del modelo causal, o si en algunas ocasiones sólo recuperan parte de la evidencia.

Por último, la investigación pretende determinar si el funcionamiento del mecanismo de predicción de acciones requiere necesariamente de la aplicación del principio de racionalidad, tal como lo proponen Csibra y Gergely. Dado que las acciones humanas no son en sí mismas evidentemente racionales, sino que adoptan este carácter únicamente en relación con la “función de utilidad” del agente, entonces es plausible hipotetizar que un mecanismo de dominio general como el modelo de Redes Bayesianas, permite hacer inferencias psicológicas incluso en circunstancias en donde el comportamiento observado no es claramente racional.

### Descripción formal del modelo de anticipación de acciones

Se asume que en toda situación en donde una persona trata de anticipar la acción de un agente observado, está implicada la participación de cuatro variables: la acción (a), la meta (m), las restricciones de la situación (w) y la situación o posición espacial a partir de la cual se inicia el movimiento (s). El modelo probabilístico de este esquema teleológico adoptaría la forma presentada en la Figura 1, en donde se observa que la acción mantiene una relación de dependencia condicional con los otros tres nodos. Adicional a estos elementos estructurales, en la Figura 1 se muestra el nodo ( $\epsilon$ ), que indica el parámetro que controla los cambios espontáneos e impredecibles en la probabilidad de ocurrencia de las acciones del agente.

FIGURA 1  
Modelo causal 1 de predicción de acciones.



Fuente: elaboración propia.

Como se muestra en la Figura 1, es posible estimar la probabilidad de las acciones de la siguiente manera:

$$P(a/m,w,s) \quad (1)$$

De acuerdo con el Teorema de Bayes, el cálculo de la probabilidad posterior se encuentra definido en función de la probabilidad previa de cada una de las acciones y la verosimilitud de los datos.

$$\text{Probabilidad previa: } P(a) \quad (2)$$

$$\text{Verosimilitud: } P(m,w,s/a) \quad (3)$$

El denominador es calculado sumando la probabilidad de todas las hipótesis posibles implicadas en la tarea.

$$\text{Normalización: } \sum_{ax \in A} P(m,w,s/ax)P(ax) \quad (4)$$

La ecuación derivada del teorema de Bayes para la anticipación de acciones es entonces la siguiente:

$$P(a/m,w,s) = \frac{P(m,w,s/a)P(a)}{\sum_{ax \in A} P(m,w,s/ax)P(ax)} \quad (5)$$

La ecuación 5 especifica, por lo tanto, la manera como la probabilidad de ocurrencia de cierta

acción se actualiza mediante los cambios en la probabilidad de la meta perseguida, las restricciones del ambiente y la situación inicial de la que parte el agente.

## Experimento 1

Mediante el primer experimento se buscó poner a prueba la validez del modelo bayesiano (sintetizado en la ecuación 5) para dar cuenta del proceso de anticipación de acciones en situaciones de alta incertidumbre. A diferencia del estudio de Baker et al. (2007) en donde sólo dos de las variables fueron probabilísticas (la meta y la acción), en esta investigación se pretendió utilizar una situación más compleja; todas las cuatro variables del modelo presentaron pesos probabilísticos diferentes. Buscamos, de esta manera, determinar si las personas pueden estimar la probabilidad de las acciones de un agente condicionadas a más de dos variables. Se trabajó con población adulta para obtener mediciones más exactas de las que se podrían obtener con niños.

Se utilizaron dos condiciones experimentales. En la primera, denominada *predicción incondicional (PIN)*, se pidió a las personas que hicieran la predicción de la siguiente acción de un agente antes de observar la situación inicial de la que partía para alcanzar su meta. La segunda condición, *predicción condicional (PC)*, se pidió a las personas que hicieran la predicción sólo después de haber observado la situación inicial de la que el agente partía. Si el modelo bayesiano es correcto y las personas representan los parámetros de todo el esquema teleológico, entonces no deberían observarse diferencias de desempeño entre ambas condiciones.

## Metodología

### Participantes

18 estudiantes de la Universidad del Valle, Colombia, participaron en este experimento. 12 eran mujeres y 6 hombres. Sus edades oscilaban entre 17 y 22 años ( $M = 19$ ). La mitad de los participantes fueron aleatoriamente asignados a la condición PIN y la otra mitad a la condición PC.

### Tarea

La tarea fue diseñada en un formato virtual mediante el programa de animación Macromedia Flash 8.0. Como se muestra en la Figura 2, se construyó un pequeño laberinto con dos círculos concéntricos, que simulaban ser muros. Ambos círculos constaban de cuatro entradas. Las entradas del círculo externo representaban la situación inicial (s) de la que partía el agente; mientras que las entradas del círculo interno representaban las acciones (a) que el agente podía ejecutar para alcanzar la meta perseguida. En el centro del laberinto, se ubicaron dos pequeñas barras de colores diferentes que simulaban ser fuentes de comida, estas barras representaban las metas (m) del agente.

Por otra parte, las restricciones (w) de la situación se representaron mediante la aparición de un bloque rojo en cualquiera de las cuatro entradas internas. Este bloque podía impedir la entrada del agente al centro del laberinto. El agente fue representado mediante un pequeño círculo azul que podía desplazarse libremente dentro del laberinto. En la parte externa del laberinto, se colocó una cubierta blanca que simulaba ser una capa de nubes alrededor del laberinto y que impedía saber por dónde entraría el agente en cada escena presentada. En las cuatro entradas del círculo externo, se escribieron cuatro letras diferentes (A, B, C, D), para facilitar su identificación, cuando las personas hicieran sus predicciones.

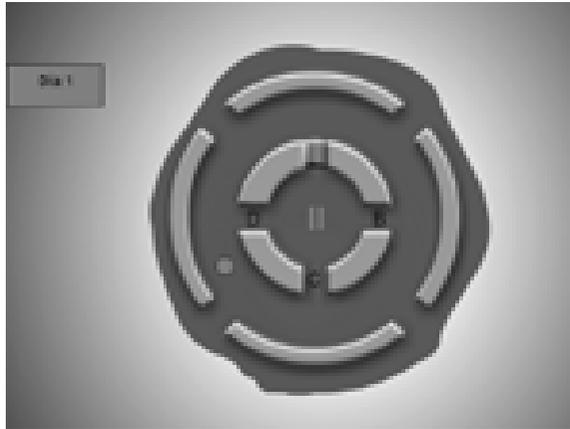
Dado que había cuatro posibles estados de la situación inicial, cuatro estados de las restricciones, dos metas y cuatro posibles acciones, el espacio total del problema estaba constituido por 128 posibles acciones o trayectorias seguidas por el agente.

A todas las personas, en ambas condiciones, se les presentó un total de 30 escenas (con una duración promedio de 7 s por escena), las cuales se representaron mediante días en la parte superior izquierda de la pantalla. Para construir cada una de las escenas se adoptó el criterio del principio de racionalidad, es decir, que el agente trataría de alcanzar su meta siguiendo la trayectoria más corta posible, esto garantizó una estructura de dependencia condicional entre todos los elementos del

esquema teleológico. Sin embargo, a pesar de esto, la ejecución de la acción racional se parametrizó con un  $\epsilon$  de 0,22. Lo que indica que en algunas ocasiones el agente se desviaba del comportamiento racional.

**FIGURA 2**

Escena de la tarea de predicción de acciones.



Fuente: elaboración propia.

### Procedimiento

La tarea fue presentada en un computador portátil con una pantalla de 15,4 pulgadas. Se comenzó con una fase de familiarización, en donde se explicó lo que cada parte de la escena representaba. El agente se presentó como una especie de extraterrestre que entraba al laberinto buscando comida. Se explicó que en cada ocasión sólo podía consumir una de las dos fuentes de alimento. La consigna de la tarea se presentó de la siguiente manera: “usted debe tratar de predecir, de la manera más exacta posible, por dónde piensa que el extraterrestre entrará durante la siguiente escena”. Se explicó también que una vez elegida una entrada debía asignar un número de 1 a 9 a las restantes entradas, de acuerdo con qué tan probable era que el extraterrestre las utilizara. Al lado del computador se colocó una hoja en donde se presentaba la escala de 1 a 9, en donde 9 era igualmente probable (en relación con la entrada elegida inicialmente), 5 la mitad de probable, y 1 absolutamente improbable.

Después de la familiarización se presentaron cuatro escenas de práctica para aclarar dudas y

observar si la persona había comprendido correctamente la consigna de la tarea. La presentación de las escenas estaba bajo control del propio participante, de tal manera que para poner al extraterrestre en movimiento debía oprimir las teclas  $\text{ctrl} + \text{enter}$ .

Una vez comprendida la tarea se pasó a la fase de prueba, en donde los participantes realizaron sus predicciones. En las cuatro primeras escenas, se decía a las personas que tan sólo observarían el comportamiento del extraterrestre y, a partir de la escena cinco, comenzarían con las predicciones. Las respuestas se registraron directamente en rejillas.

### Resultados y Discusión

Las respuestas numéricas de los participantes fueron previamente estandarizadas de tal manera que la suma de las cuatro predicciones (una por cada entrada) siempre diera 1. El conjunto de predicciones resultantes fueron correlacionadas con las predicciones arrojadas por el modelo bayesiano de anticipación de acciones, el cual arrojó un total de 65 predicciones. Se utilizó la prueba de correlación de Pearson dada la naturaleza numérica de los datos.

En la condición PIN no se obtuvo una correlación significativa ( $r = 0,13$ ), lo que indica que el comportamiento predictivo de los participantes no fue capturado por el modelo bayesiano. Se corrió una segunda prueba de correlación entre todos los participantes en esta condición, para determinar si hubo un patrón de respuestas, aunque éste no fuera bayesiano. Los resultados no mostraron correlaciones significativas entre las respuestas de los participantes (las correlaciones oscilaron entre  $r = 0,07$  y  $r = 0,2$ ). Esto sugiere que los desempeños de cada persona ante la tarea, fueron notoriamente diferentes entre sí.

Los resultados en la condición PC, por el contrario, mostraron una correlación significativa a un  $p < .01$  ( $r = 0,66$ ). Aunque la correlación no fue muy alta, este resultado muestra que las predicciones de los participantes, se adecuan bien a las predicciones del modelo bayesiano. Para tratar de determinar si la correlación obtenida fue estable a

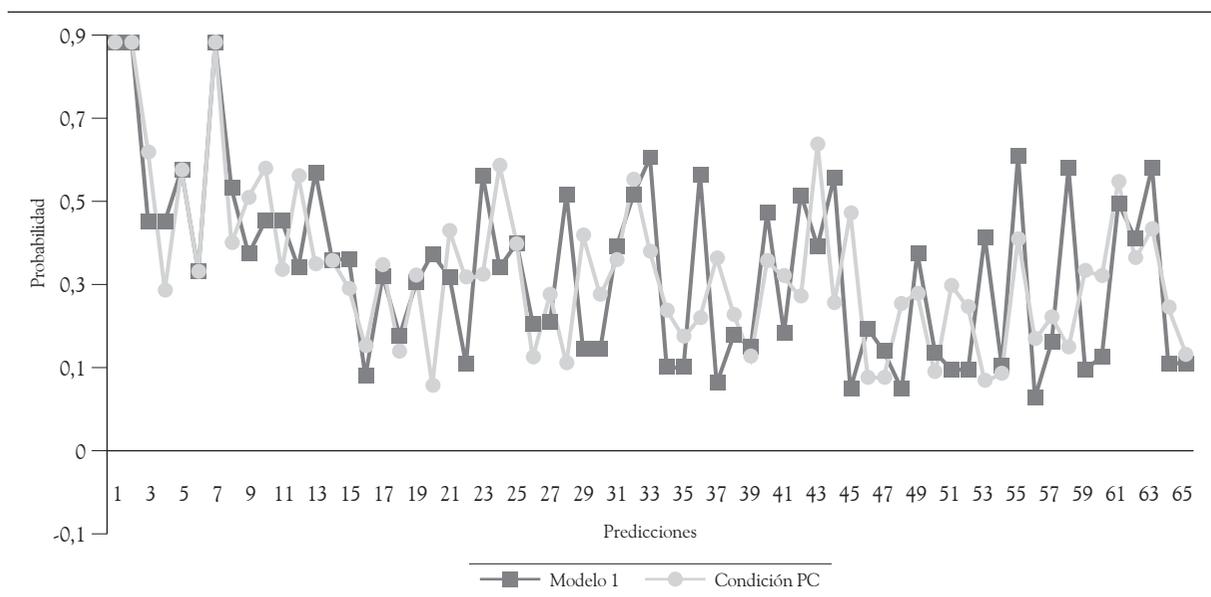
lo largo de todas las 65 predicciones, se corrió una segunda prueba dividiendo la muestra de datos en tres partes, con 22, 22 y 21 predicciones, respectivamente. La correlación de la primera sección de datos fue significativa a un  $p < .01$  ( $r = 0,87$ ), mientras que las correlaciones de las últimas dos secciones no fueron significativas ( $r = 0,21$  para la segunda sección, y  $r = 0,33$  para la tercera sección).

Los resultados obtenidos en ambas condiciones no apoyan la hipótesis originalmente propuesta, según la cual en ambas condiciones experimentales debería encontrarse una correlación significativa con el modelo bayesiano. Una primera explicación de estos resultados es que la condición PIN era notoriamente más complicada de procesar para generar una predicción aproximada, en relación con la condición PC. Sin embargo, en términos estrictos ambas condiciones presentan la misma complejidad computacional. A pesar que en la condición PC se indica al participante la situación inicial a partir de la cual realizar la predicción, esto no disminuye la dificultad del problema. De igual manera, es necesario considerar la distribución de probabilidad total del sistema, para generar una respuesta aproximada al modelo bayesiano.

Una segunda explicación de la diferencia en los desempeños, radica en que en la condición PC los participantes codificaron de manera prioritaria la información referente a la probabilidad condicional entre la situación inicial y la acción del agente. Esto pudo haber ocurrido en la medida en que la predicción solicitada destacaba claramente estas dos variables del esquema teleológico (“si el agente entra por *s2* ¿cuál es la acción (a) más probable que realizará?”). Por el contrario, en la condición PIN, la predicción requería de todas las fuentes de información del modelo al mismo tiempo. En definitiva, esto querría decir que la parametrización sucesiva del modelo en la condición PIN resultó ser excesiva para las personas, y que la solución en la condición PC se debió a una reducción del espacio de búsqueda a sólo dos variables.

Para poner a prueba la segunda interpretación de los datos, se elaboró un nuevo modelo bayesiano, de acuerdo con el cual el proceso de predicción de acciones estaría condicionado únicamente al valor cambiante de la probabilidad de la situación inicial. Es decir, que las predicciones deberían adecuarse al patrón mostrado por la ecuación 6:

FIGURA 3  
Correlación resultados condición PC con modelo causal 1.



Fuente: elaboración propia.

$$P(a/s) = \frac{P(s/a)P(a)}{\sum_{ax \in A} P(s/ax)P(ax)} \quad (6)$$

El resultado de la correlación de las respuestas de los participantes en la condición PC con el segundo modelo bayesiano mostró una correlación significativa al  $p < .01$  de  $r = 0,76$ , más alta que la obtenida con el primer modelo. Sin embargo, al dividir la muestra de datos en tres cortes, con 19 puntuaciones cada una, se obtuvieron las siguientes correlaciones (en esta ocasión el modelo sólo hizo 57 predicciones): la correlación para la primera sección de datos fue significativa a un  $p < .05$  ( $r = 0,54$ ), mientras que las correlaciones con los cortes 2 y 3 fueron significativas a un  $p < .01$  ( $r = 0,78$  y  $r = 0,86$  respectivamente).

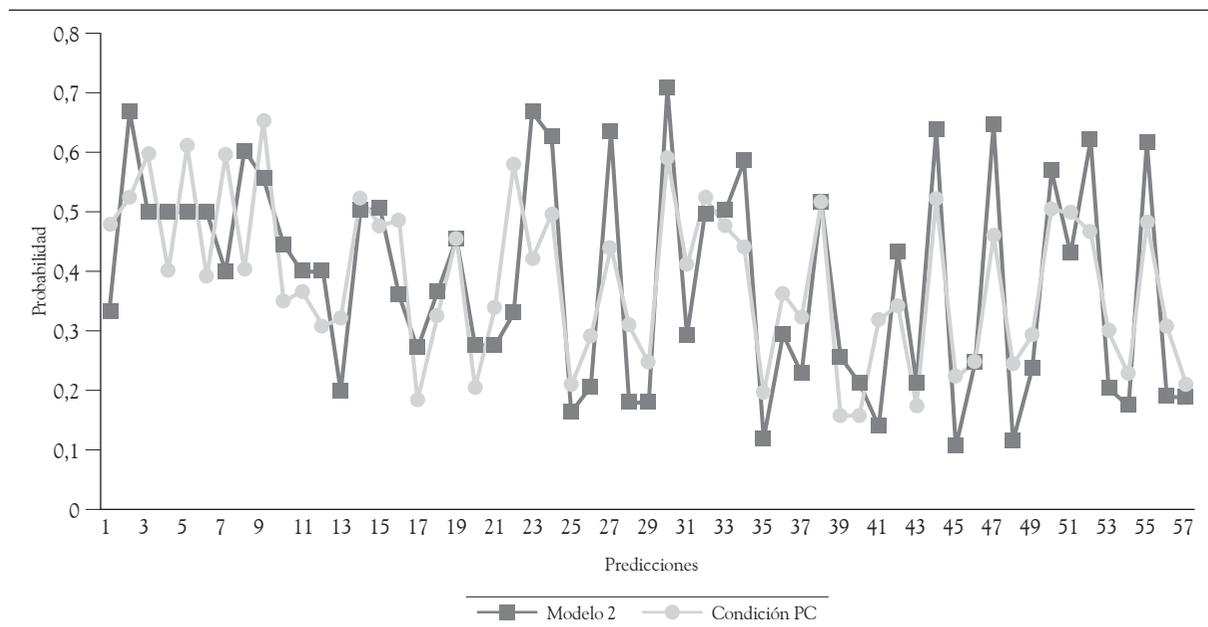
En conjunto, todos los resultados anteriores muestran un panorama más complejo al inicialmente pensado. Al parecer, el desempeño de los participantes a lo largo de las 30 observaciones varió en función del modelo bayesiano utilizado. Durante la primera parte de la tarea emplearon el modelo expuesto en la ecuación 5 (con un  $r$  de 0,87), y en las últimas dos partes utilizaron el mo-

delo presentado en la ecuación 6 (con un  $r$  de 0,78 para la segunda parte y un  $r$  de 0,86 para la tercera). Posiblemente, a medida que la tarea avanzó las personas variaron su centro de atención, y pasaron de hacer predicciones a partir de un modelo de cuatro variables a hacer predicciones, a partir de un modelo con sólo dos variables.

Estos resultados apoyan parcialmente la segunda interpretación propuesta para explicar la diferencia de resultados entre las dos condiciones experimentales utilizadas. En efecto, durante la mayor parte de la tarea las personas anticipan las acciones del agente a partir de la información sobre la situación inicial; sin embargo, el hecho que en la primera sección de datos hayan empleado el modelo completo de cuatro variables, sugiere que el fracaso para anticipar las acciones en la condición PIN no se debió a la dificultad intrínseca de la tarea.

Una tercera forma de interpretar los resultados, es afirmar que las diferencias en los desempeños no provienen sólo de dificultades *computacionales*, sino, primordialmente, de diferencias en la *representación* del modelo causal. La predicción de

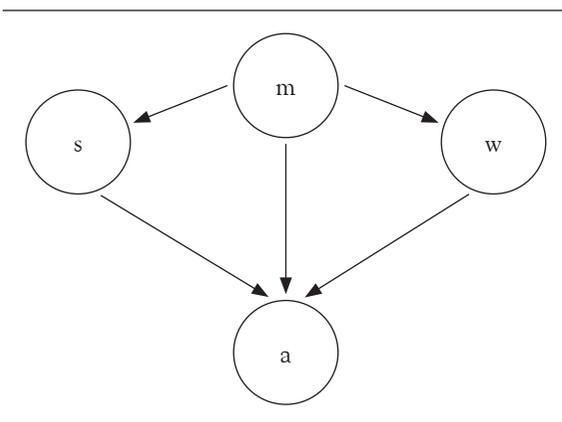
FIGURA 4  
Correlación resultados condición PC con modelo causal 2.



Fuente: elaboración propia.

acciones en la condición PIN exige la utilización del modelo causal presentado en la Figura 1; pero la predicción en la condición PC puede realizarse también mediante el modelo causal de la Figura 5. En apariencia más complicado, tiene la ventaja de dividir el procesamiento de la información en dos etapas. Primero, se estima la probabilidad de las restricciones ( $w$ ) y de la meta ( $m$ ) condicionadas a la situación inicial ( $s$ ); esto ocurre cuando se le presenta a la persona la posición de la que parte el agente, lo que le permite predecir cuál será la meta más probable que intentará alcanzar. La segunda etapa consiste en estimar la probabilidad de las acciones ( $a$ ) condicional a ( $w$ ), ( $m$ ) y ( $s$ ). Como se demostró atrás, las personas realizan la predicción a partir de todas las variables y no únicamente de ( $s$ ) (al menos en la primera parte de la tarea), esto exige entonces que los efectos de la situación inicial sobre las acciones estén mediados tanto por la meta como por las restricciones, tal como se observa en el Gráfico 5.

FIGURA 5  
Modelo causal 2 de la predicción de acciones.



Fuente: elaboración propia.

De acuerdo con lo anterior, la facilidad relativa que encontraron los participantes para hacer predicciones condicionales consistió en que generaron una representación mejor organizada de la información. La primera etapa de la inferencia representa únicamente la probabilidad condicional entre  $m/s$  y  $w/s$ . Las personas saben de dónde parte el agente y cuál es la meta más probable. A

partir de este conocimiento determinan cuál es la acción más probable.

Una pregunta derivada a partir de la interpretación anterior es si la dificultad que tienen las personas para realizar predicciones incondicionadas puede eliminarse si se disminuye el número de variables del esquema teleológico. Si la respuesta a esta pregunta es positiva, entonces indicaría que la dificultad estaría relacionada también con la cantidad de información que deben manejar las personas y no con el tipo de representación involucrada, como hemos propuesto.

## Experimento 2

El segundo experimento tuvo un doble propósito. En primer lugar, pretendió resolver la pregunta atrás formulada, para hallar nueva evidencia a favor o en contra de la validez del modelo causal 2 de la predicción de acciones. En segundo lugar, pretendió determinar si la predicción de acciones requiere necesariamente de la aplicación del principio de racionalidad para funcionar, o si la inferencia bayesiana opera incluso bajo circunstancias de comportamiento no racional.

Se emplearon dos condiciones experimentales. En la condición de “predicción irracional” (PI) se presentó una secuencia de comportamientos en donde el agente tendía a violar consistentemente el principio de racionalidad; es decir, que no seguía la trayectoria más corta entre la situación inicial y la meta buscada. En la segunda condición, “predicción reducida” (PR), se presentó una tarea de predicción con sólo dos variables probabilísticas, la restricción de entrada ( $w$ ) y la acción del agente ( $a$ ).

## Metodología

### Participantes

18 estudiantes de la Universidad del Valle, Colombia, participaron en este experimento. Fueron diferentes a los que participaron en el experimento anterior. 11 eran mujeres y 7 hombres. Sus edades oscilaron entre 17 y 23 años ( $M = 20$ ). La mitad de

los participantes fueron aleatoriamente asignados a la condición PI y la otra mitad a la condición PR.

### *Tareas*

Las tareas empleadas fueron similares a las del experimento anterior. Para la condición PI, se utilizó un laberinto con las mismas características físicas que en la condición de “predicción condicional”, la única diferencia consistió en que las trayectorias observadas no mostraban un comportamiento racional. Para conseguir este efecto, en algunas ocasiones el agente elegía la entrada más lejana en relación con su posición inicial, y en otras ocasiones elegía la entrada más cercana, pero seguía una ruta muy larga para alcanzarla; por ejemplo, daba un giro completo al laberinto antes de entrar. La acción del agente se parametrizó con un  $\epsilon$  de 0,2. Esto indica que en algunas trayectorias el agente se desviaba de su conducta irracional y realizaba acciones racionales.

En la condición PR se utilizó un laberinto con una sola meta, es decir, que se indicó a las personas que el extraterrestre perseguía una única fuente de alimento. El círculo externo del laberinto tenía una sola entrada, de tal manera que la situación inicial ( $s$ ) fue la misma durante toda la tarea. El círculo interno tenía cuatro entradas, al igual que en el experimento anterior. En esta ocasión, se utilizaron dos bloques rojos que tapaban al mismo tiempo dos de las cuatro entradas del círculo interno. Esta variación se introdujo para aumentar los posibles estados de la variable  $w$ . Dado que había cuatro acciones posibles y seis estados de las restricciones (los dos bloques rojos se combinan en seis posiciones posibles), el espacio total del problema estaba constituido por 24 trayectorias posibles. La acción del agente se parametrizó con un  $\epsilon$  de 0,2. Esto quiere decir que en la mayoría de las ocasiones el comportamiento del agente fue totalmente racional, excepto en el 20 % de las escenas.

### *Procedimiento*

En la condición PI se mostraron 30 escenas a cada participante, mientras que en la condición PR sólo

se presentaron 21 escenas. La razón por la cual se presentaron menos escenas en la condición PR fue que el tamaño pequeño del espacio del problema aumentaba la probabilidad de repetir escenas; para evitar que las inferencias de los participantes estuvieran basadas en el recuerdo de trayectorias específicas, se optó por reducir el número de escenas presentadas. En todos los demás aspectos relevantes el procedimiento de aplicación en ambas condiciones fue idéntico al utilizado en el experimento anterior.

## **Resultados y Discusión**

Las respuestas de los participantes fueron previamente estandarizadas para que la suma de las predicciones numéricas diera siempre 1. La aplicación de la ecuación 5 para la tarea empleada en la condición PI arrojó un total de 75 predicciones. Para analizar los datos en la condición PR se utilizó la ecuación 6, en donde la variable ( $s$ ) se reemplazó por ( $w$ ); en esta condición la ecuación arrojó 60 predicciones. Se utilizó nuevamente la prueba de correlación de Pearson.

En la condición PI se obtuvo una correlación significativa con un  $p < .01$  ( $r = 0,6$ ). Este resultado indica que las predicciones de los participantes se ajustaron al modelo bayesiano. La muestra de datos se dividió en tres partes, de 25 puntuaciones cada una, para determinar la estabilidad de la correlación. Los primeros 25 datos correlacionaron significativamente con un  $p < .01$  ( $r = 0,75$ ), mientras que los otros dos grupos de datos no mostraron una correlación significativa;  $r = 0,2$  para el segundo grupo y  $r = 0,1$  para el tercero. Estos resultados replican los obtenidos en la condición PC del experimento anterior, aunque con correlaciones ligeramente más bajas. Posiblemente la percepción de un comportamiento irracional interfirió en el proceso de inferencia psicológica.

En la condición PR no se obtuvo una correlación significativa ( $r = 0,1$ ). Cuando la muestra de datos se dividió en tres grupos tampoco se obtuvieron correlaciones significativas. Estos resultados replican los obtenidos en la condición PIN del experimento anterior, e indican que las respuestas

de los sujetos no se adecuaron al modelo bayesiano propuesto en la ecuación 6.

En conjunto, los resultados de este experimento muestran cómo ante una tarea de baja dificultad (en términos de la amplitud del espacio del problema y el número de variables del modelo causal) como la utilizada en la condición PR, los participantes no parecen guiar sus predicciones a partir de información probabilística; pero, en contraste, ante una tarea de mayor dificultad, como la presentada en la condición PI, los participantes sí consiguen recuperar la evidencia de la probabilidad condicional entre las variables.

Estos resultados sugieren, en primer lugar, que el mecanismo de inferencia bayesiana funciona incluso en circunstancias en donde el comportamiento de los agentes no es racional. Esto soporta la idea según la cuál bajo condiciones de incertidumbre la anticipación de las acciones no debe guiarse únicamente mediante la aplicación cualitativa del principio de racionalidad, sino que puede basarse además en un mecanismo de dominio general como la inferencia bayesiana. En segundo lugar, los resultados sugieren nuevamente que la diferencia en los desempeños entre la predicción condicional y la predicción incondicional, no yace en la dificultad de la tarea, sino en la manera como se organiza el conocimiento en cada caso. Incluso cuando la predicción incondicionada se realizó con sólo dos variables probabilísticas, las personas no fueron capaces de hacer uso de la información suministrada.

## Discusión General

Los anteriores experimentos muestran la coherencia de la inferencia psicológica que realizan las personas en condiciones de incertidumbre. En específico, la anticipación de acciones parece adecuarse a un proceso de inferencia bayesiana. Esto indica dos cosas; en primer lugar, que la inferencia psicológica, al igual que la inferencia causal, opera a partir de la recuperación del modelo causal de la situación propuesta; y en segundo lugar, que las personas son capaces de actualizar la probabilidad condicional de los parámetros del esquema teleo-

lógico, mediante un proceso análogo al dictado por el Teorema de Bayes.

La evidencia más fuerte a favor del rol determinante de los modelos causales en la inferencia psicológica se obtuvo en la diferencia de desempeños entre la predicción condicionada y la predicción incondicionada. Ambas inferencias pueden ser resueltas mediante una misma ecuación bayesiana, con el mismo grado de dificultad; sin embargo, el análisis de los datos sugiere que en cada caso las personas utilizaron modelos causales diferenciados. Esto provocó que un grupo de participantes realizara predicciones más o menos correctas (predicción condicionada), mientras que otro grupo no logró utilizar la información mostrada en los videos (predicción incondicionada).

La incidencia de los modelos causales en el razonamiento ha sido notada en varias investigaciones. Waldmann y Hagmayer (2005), por ejemplo, muestran cómo ante una pequeña reorganización inducida de un modelo causal (con el mismo tipo y número de nodos), las personas generan inferencias drásticamente diferentes; efecto similar al que se obtuvo en la actual investigación. Esta similitud además brinda soporte adicional a la teoría de las “redes jerárquicas” propuesta por Tenenbaum et al. (2006), según la cual los modelos probabilísticos bayesianos son empleados más allá del dominio de la causalidad y permiten explicar también la inferencia biológica y psicológica, como en el caso actual.

Más allá de la construcción de los modelos causales, los resultados indican la capacidad de las personas para “estimar los parámetros” del modelo construido y de actualizar sus valores a través de observaciones sucesivas. No obstante, tanto en la condición PC del primer experimento como en la condición PI del segundo, se observa que el proceso de actualización no es constante ni homogéneo. En ambas condiciones se obtuvo una alta correlación sólo en la primera sección de las predicciones; más allá, la correlación con el modelo bayesiano de la ecuación 5 decayó drásticamente, al parecer a favor de un segundo modelo (ecuación 6). Este resultado no había sido obtenido en estudios anteriores.

En la investigación de Baker et al. (2007) se mostró cómo la decisión racional de predecir la

meta de un agente estaba mediada por un análisis bayesiano, lo que justifica a algunos investigadores para hablar de un “análisis racional Bayesiano”, tema de vieja data en la literatura económica (Hansson, 1994). Los resultados mostrados en la actual investigación, sin embargo, sugieren que el mecanismo de inferencia bayesiano puede servir incluso para inferir la “función de utilidad” del agente, es decir, para detectar patrones y determinar lo que hace racional su comportamiento en un momento dado. Como Bonawitz, Griffiths y Schulz (2006) lo hacen notar, la importancia de poseer un mecanismo de inferencia de dominio general, como las Redes Bayesianas, radica en que permite al sistema cognitivo superar las restricciones propias de un cuerpo de conocimiento ya dado, mediante la inferencia de nuevos principios generales, a partir de eventos y regularidades observadas en el mundo.

## Referencias

- Baker, C. L., Tenenbaum, J. B. & Saxe, R. R. (2006). Bayesian models of human action understanding. *In Advances in Neural Information Processing Systems*, 18, 99-106.
- Baker, C. L., Tenenbaum, J. B. & Saxe, R. R. (2007). *Goal inference as inverse planning*. Recuperado el 28 de septiembre, 2008, de <http://web.mit.edu/clbaker/www/papers/cogsci2007.pdf>
- Bonawitz, E. B., Griffiths, T. L. & Schulz, L. E. (2006). Modeling cross-domain causal learning in preschoolers as Bayesian inference. En R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 89-94). Mahwah, NJ: Lawrence Erlbaum Associates.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society, London B*, 358, 447-458
- Csibra, G. & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of action in humans. *Acta Psychologica*, 124, 60-78.
- Csibra, G., Bíró, S., Koós, O. & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, 27, 111-133.
- Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *TRENDS in Cognitive Sciences*, 7(7), 287-292.
- Gergely, G., Nadasdy, Z., Csibra, G. & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-193.
- Glymour, C. (2003). Learning, prediction and causal bayes nets. *TRENDS in Cognitive Sciences*, 7(1), 43-48.
- Gopnik, A. & Schulz, L. (2004). Mechanism of theory formation in young children. *TRENDS in Cognitive Sciences*, 8(8), 43-48.
- Gopnik, A. & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10(3), 281-287.
- Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T. & Danks, D. (2004). A theory of learning in children: Causal maps and bayes nets. *Psychological Review*, 111 (1), 1-30.
- Griffiths, T. & Tenenbaum, J. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 334-384.
- Griffiths, T., Kemp, C. & Tenenbaum, J. (2008). *Bayesian Models*. Manuscrito no publicado.
- Hansson, S. O. (1994). *Decision theory: A brief introduction*. Stockholm: Royal Institute of Technology.
- Körding, K. & Wolpert, D. (2006). Bayesian decision theory in sensorimotor control. *TRENDS in Cognitive Sciences*, 10(7), 319-326.
- Meder, B., Hagmayer, Y. & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review*, 15(1), 75-80.
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838-850.
- Rao, R., Shon, A. & Meltzoff, A. (2004). A Bayesian model of imitation in infants and robots. En K. Dautenhahn & C. Nehaniv (Eds.), *Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions* (pp. 217-248), Cambridge: Cambridge University Press.

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *TRENDS in Cognitive Sciences*, 10(7), 309-318.

Waldmann, M. R. & Hagmayer, Y. (2005). Seeing versus doing: two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31 (2), 216-227.

