

Sesgos de la calibración en tareas de razonamiento lógico

Calibration biases in logical reasoning tasks

Guillermo Macbeth¹, Alfredo O. López Alonso, Eugenia Razumiejczyk
Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Universidad del Salvador, Argentina

Rodrigo A. Sosa, Carolina I. Pereyra

Universidad del Salvador, Argentina

Humberto Fernández

Universidad Abierta Interamericana, Argentina

(Recepción: Octubre 2009 – Aceptación: Diciembre 2009)

Resumen

Se presenta un estudio experimental sobre calibración en tareas deductivas. Se define la calibración como la relación de convergencia o divergencia empírica entre el éxito objetivo y el éxito subjetivo. Cuando el primero predomina sobre el segundo se presenta el sesgo de subconfianza. La hipótesis de trabajo de este estudio sostiene que la calibración en tareas de razonamiento lógico depende de la forma de las proposiciones que presente el experimento. Se distingue la afirmación de la negación en el procesamiento cognitivo de las proposiciones compuestas. Los resultados sugieren que las formas proposicionales del mismo signo o monótonas generan subconfianza. Se propone explicar este fenómeno por la activación de un heurístico de monotonía que produce una ilusión de simplicidad responsable de la generación del sesgo. Se discuten estos hallazgos en el contexto de los modelos metacognitivos de la calibración.

Palabras clave: Calibración, sesgos, razonamiento, heurístico de monotonía, ilusión de simplicidad.

Abstract

The aim of this contribution is to present an experimental study about calibration in deductive reasoning tasks. Calibration is defined as the empirical convergence or divergence between the objective and the subjective success. The underconfidence bias is understood as the dominance of the former over the latter. The hypothesis of this study states that the form of the propositions presented in the experiment is critical for calibration phenomena. Affirmative and negative propositions are distinguished in their cognitive processing. Results suggests that monotonous compound propositions are prone to underconfidence. An heuristic approach to this phenomenon is proposed. The activation of a monotony heuristic would produce an illusion of simplicity that generates the calibration bias. These evidence is analysed in the context of the metacognitive modeling of calibration phenomena.

Key words: Calibration, biases, reasoning, monotony heuristic, simplicity illusion.

1 Correspondencia: Guillermo Macbeth. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Instituto de Investigaciones Psicológicas de la Universidad del Salvador (IIPUS). Marcelo T. de Alvear 1314 (C1058AAV), 1º piso, Ciudad Autónoma de Buenos Aires, Argentina. Teléfono: (+5411)-4813-3404 E-mail: guillermo.macbeth@mail.salvador.edu.ar / guimacbeth@hotmail.com

Distorsiones de la calibración

Hace más de cuatro décadas propuso Oskamp (1965) un modelo matemático general para el estudio de los fenómenos psicológicos de la calibración. En este modelo, la calibración o confianza C puede ser entendida como una relación de convergencia o divergencia empírica entre el éxito objetivo O y el éxito subjetivo E alcanzado por una persona en la realización de una colección de n cantidad de tareas (Gigerenzer, Hoffrage & Kleinböling, 1991; Koehler, Brenner & Griffin, 2002). El éxito objetivo remite al rendimiento propiamente observado. El éxito subjetivo, en cambio, se refiere a la creencia o estimación personal que posee la persona acerca de su propio desempeño. La relación entre ambas variables permite definir tres fenómenos básicos de la calibración (Kahneman & Tversky, 1982; Lichtenstein, Fischhoff & Phillips, 1982), como se indica en la Ecuación 1

$$C = \sum_{i=1}^n E_i - O_i = \begin{cases} (E > O) = (C+) = \text{sesgo de sobreconfianza} \\ (E < O) = (C-) = \text{sesgo de subconfianza} \\ (E = O) = (C = \text{cero}) = \text{ausencia de sesgos} \end{cases} \quad \forall n \in \mathbb{N} \quad (1)$$

Sólo cuando el éxito subjetivo y el éxito objetivo coinciden se presenta una buena calibración o ausencia de sesgos (Merkle & Van Zandt, 2006). Cuando el primero es mayor que el segundo se observa el sesgo de sobreconfianza. En el caso contrario, la distorsión hallada es conocida como sesgo de subconfianza (Simmons & Nelson, 2006).

En estudios recientes (Macbeth, 2009a; Macbeth & Fernández, 2008) se ha propuesto un nuevo modelo matemático coherente con la definición de Oskamp (1965) que define la calibración como una función polinómica en reales con $x \in O$ como dominio e $y \in E$ como codominio. Este modelo asume que: i) la ausencia de sesgos o buena calibración es una función identidad $B_{(x)} = x$, ii) la sobreconfianza depende de alguna función $C_{(x)}^+$ con imagen siempre mayor que $B_{(x)}$ y, iii) la subconfianza se comporta como una relación funcional desconocida $C_{(x)}^-$ con imagen siempre menor que $B_{(x)}$. Sobre la base de estos tres supuestos se redefinieron los sesgos de la calibración según las Ecuaciones 2 y 3, en el intervalo real $[a, b]$ determinado por los extremos inferior y superior de la escala que se aplique para medir O y E .

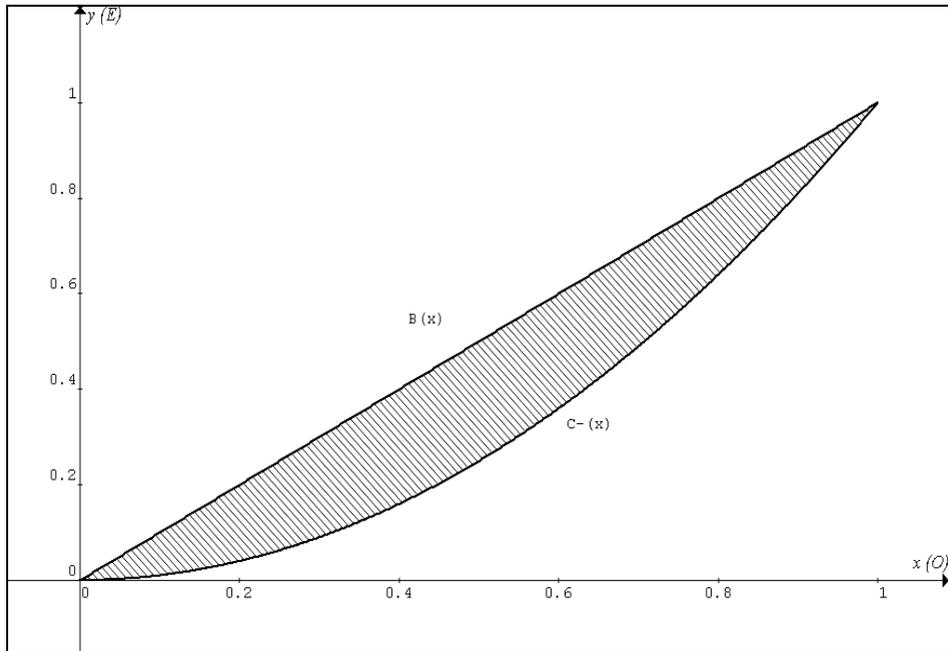
$$\text{Sesgo de sobreconfianza} = \int_a^b (C_{(x)}^+ - B_{(x)}) dx \quad (2)$$

$$\text{Sesgo de subconfianza} = \int_a^b (B_{(x)} - C_{(x)}^-) dx \quad (3)$$

Si bien las funciones $C_{(x)}^+$ y $C_{(x)}^-$ son teóricas y por ello generales, se propone en la Figura 1 una representación gráfica concreta del modelo con el propósito de ilustrar una de las posibles relaciones observables entre la buena calibración y el sesgo de subconfianza.

Figura 1

El sesgo de subconfianza como integral definida



Nota: se propone definir el sesgo de subconfianza como el área comprendida entre las funciones

$B(x)$ y $C-(x)$, entendida como la integral definida en este ejemplo para el intervalo de $x=[0,1]$.

Los puntos del área sombreada de la Figura 1 corresponden a casos posibles de subconfianza. Se empleó para este ejemplo la aproximación cuadrática $C-(x) = x^2$ para modelar el comportamiento empírico del sesgo, pero también se podrían emplear otros ajustes de acuerdo a la forma de la distribución que presenten las observaciones. Dos ventajas importantes que ofrece esta extensión del modelo de Oskamp (1965) consisten en la posibilidad de: i) aplicar cualquier escala en reales para medir el éxito objetivo y el éxito subjetivo y, ii) cuantificar la magnitud del sesgo mediante el cálculo del área comprendida entre la buena calibración y alguna función que aproxime el comportamiento empírico de la sobreconfianza o la subconfianza. La cuantificación del sesgo puede realizarse mediante el método de Barrow para integración definida (Spivak, 1992) y la aproximación de la función empírica puede ajustarse mediante mínimos cuadrados (Grossman, 2007). En estudios recientes también se ha empleado el tamaño del efecto asociado a la prueba de hipótesis para la diferencia entre el éxito objetivo y el éxito subjetivo como estimador de la magnitud del sesgo (Macbeth, 2009a; Macbeth & Fernández, 2008; Macbeth, Ledesma, Razumiejczyk, Cortada de Kohan & López Alonso, 2008; Macbeth & Morán, 2009; Macbeth & Razumiejczyk, 2008).

Mediante la aplicación de este modelo de calibración se ha estudiado el comportamiento de la sobreconfianza y la subconfianza en tareas de diversos dominios. Se han encontrado sesgos en tareas verbales (Macbeth & Cortada de Kohan, 2008), matemáticas (Macbeth, Razumiejczyk & Cortada de Kohan, 2006), económicas (Camerer & Lovallo, 1999), médicas (Oskamp, 1965), perceptivas (Razumiejczyk, Pereyra & Macbeth, 2009) y de conocimientos generales (Macbeth, Cortada de Kohan, Razumiejczyk & López Alonso, 2006), entre otras (Macbeth & López Alonso, 2008; Svenson, 1981). En tareas de razonamiento deductivo se han realizado sólo algunos estudios (Brenner, Koehler, Liberman & Tversky, 1996; Evans, 1984). Ninguna de las investigaciones que conforman el estado actual de los conocimientos sobre el tema se ha ocupado de la relación específica entre

calibración y razonamiento lógico que se propone en el presente estudio. En un experimento previo se ha encontrado que la resolución de problemas matemáticos genera el sesgo de sobreconfianza tanto en estudiantes de tecnología, como en estudiantes de humanidades (Macbeth, 2009b). Se ha observado también que la magnitud de estos sesgos no se reduce fácilmente (Fischhoff, 1982; Gigerenzer et al., 1991). La advertencia directa, la recompensa económica y la adquisición de experticia no han logrado reducir sistemáticamente las distorsiones de la calibración (Brenner, 2003; Camerer & Lovallo, 1999; Juslin, Winman & Olsson, 2000). Sólo mediante manipulaciones experimentales centradas en el monitoreo metacognitivo se han logrado moderar e incluso eliminar los sesgos de sobreconfianza y subconfianza (Macbeth & Razumiejczyk, 2008; Wilson, Centerbar & Brekke, 2002). La metacognición se entiende en este contexto como una recursividad compleja del procesamiento de la información que se encuentra relacionada con un registro de nivel superior y permite lograr cierto control sobre algunos procesos cognitivos básicos (Fernandez-Duque & Black, 2007; Flavell, 1979; Nelson, 1996).

Calibración en tareas de razonamiento lógico

Una de las variedades del razonamiento lógico que más se ha estudiado es el razonamiento proposicional (Asensio, Cordero, García Madruga & Recio, 1990; Santamaría & Johnson-Laird, 1988). En esta línea de investigación se ha encontrado que la cualidad de las proposiciones resulta relevante para el reconocimiento de la validez de los silogismos (Dasí & Algarabel, 2003). Esta cualidad se refiere a la forma afirmativa o negativa que presenta la declaración que conforma una proposición. De esta manera, la forma afirmativa p y su negativa $\neg p$ reciben distinto procesamiento cognitivo en tareas de reconocimiento de validez inferencial. Cuando dos proposiciones simples p y q se combinan en una única proposición compuesta, se generan cuatro formas posibles de relación R entre ellas: 1) pRq , 2) $pR\neg q$, 3) $\neg pRq$ y 4) $\neg pR\neg q$. Las relaciones que más se han estudiado son los conectores lógicos de afirmación, negación, conjunción, disyunción y, sobre todo, implicación (Valiña & Vega, 1988). En el presente estudio interesa estudiar específicamente tres relaciones modales que han sido definidas como necesidad, imposibilidad y contingencia (López Alonso, 1991a). La relación de necesidad se refiere a las proposiciones compuestas que presentan la forma p tiene que ser q , la imposibilidad corresponde a p no puede ser q y la contingencia a p puede ser q (López Alonso, 1988).

Si bien no se cuenta con investigaciones previas sobre calibración y cualidad de las proposiciones en el razonamiento lógico, la hipótesis de trabajo del presente estudio afirma que la ocurrencia de sesgos en este dominio depende de la dificultad específica de las tareas experimentales. En coherencia con el estado actual de los conocimientos sobre calibración se pronostica el sesgo de sobreconfianza para tareas difíciles y el sesgo de subconfianza para tareas relativamente fáciles. Este fenómeno es conocido como *efecto difícil-fácil* y su ocurrencia se ha registrado en múltiples dominios (Gigerenzer et al., 1991; Juslin et al., 2000; Klayman, Soll, Gonzalez-Vallejo & Barlas, 1999).

Diversos estudios previos sobre coherencia del razonamiento lógico ofrecen para la presente investigación un marco teórico adecuado y un instrumento válido para la medición de estos fenómenos. El Test de Coherencia de Razonamiento (TCR) diseñado por López Alonso (1988, 1991a, 1991b, 1996, 2000) es una escala válida y confiable que posee adecuadas propiedades psicométricas para la evaluación del éxito objetivo y el éxito subjetivo en tareas de razonamiento lógico. El TCR presenta un total de 64 proposiciones compuestas que relacionan dos proposiciones simples previamente definidas por la persona evaluada. De esta manera, se emplean cuatro términos, cuya definición se solicita en la consigna. Estos términos p , q , r y s , cuyas definiciones debe dar el participante evaluado son, respectivamente: 'Una persona americana es...', 'Una persona argentina es...', 'Una persona europea es...' y 'Una persona inglesa es...'. El conteo de todas las variaciones posibles que surgen de combinar en una proposición compuesta estas cuatro proposiciones p , q , r , s y sus cuatro negaciones $\neg p$, $\neg q$, $\neg r$ y $\neg s$, tomadas de a dos, con reemplazo y de manera ordenada, genera los $8^2 = 64$ ítems que componen el TCR. La tarea principal de esta prueba consiste en indicar para cada proposición la relación modal R de necesidad, posibilidad o imposibilidad que se sigue lógicamente de las definiciones dadas como premisas por la persona evaluada (López Alonso, 1988). Debido a que la definición de las premisas es dada por la persona evaluada, el TCR permite identificar dife-

rentes figuras lógicas a partir de las cuales el experimentador puede evaluar la coherencia general del razonamiento e identificar diversas falacias. En estudios previos se ha encontrado que las figuras más típicas son las que presentan menor cantidad de errores lógicos (López Alonso, 1991a). Las figuras idiosincrásicas, por el contrario, tienden a generar mayor cantidad de errores y se encuentran asociadas a poblaciones psiquiátricas. El TCR se organiza en cuatro subescalas de 16 ítems cada una, cuya forma se indica en la Tabla 1.

Tabla 1 - Forma lógica del TRC

Subescala (ítems)	Forma de las proposiciones
I (1-16)	$\{p \vee q \vee r \vee s\} \mathcal{R} \{p \vee q \vee r \vee s\} = \{afirmación\} \mathcal{R} \{afirmación\}$
II (17-32)	$\{p \vee q \vee r \vee s\} \mathcal{R} \{\neg p \vee \neg q \vee \neg r \vee \neg s\} = \{afirmación\} \mathcal{R} \{negación\}$
III (33-48)	$\{\neg p \vee \neg q \vee \neg r \vee \neg s\} \mathcal{R} \{p \vee q \vee r \vee s\} = \{negación\} \mathcal{R} \{afirmación\}$
IV (49-64)	$\{\neg p \vee \neg q \vee \neg r \vee \neg s\} \mathcal{R} \{\neg p \vee \neg q \vee \neg r \vee \neg s\} = \{negación\} \mathcal{R} \{negación\}$

Método

Se empleó un diseño completamente aleatorizado para poner a prueba una serie de hipótesis experimentales relacionadas con el sesgo de subconfianza y su disolución en tareas de razonamiento lógico. Se determinaron dos grupos aleatorios, uno control y otro experimental. En el grupo control se evaluó el comportamiento de la calibración en las cuatro subescalas del TCR, sin manipulación. Este grupo mostró subconfianza en dos de las cuatro escalas. En el grupo experimental se empleó una estrategia metacognitiva utilizada en estudios previos para disolver el sesgo de subconfianza hallado en el grupo control.

Participantes

Participaron del experimento 152 estudiantes universitarios de grado de la Universidad de Buenos Aires, la Universidad del Salvador y la Universidad Abierta Interamericana, todas ubicadas en la ciudad de Buenos Aires, Argentina. La participación de mujeres (58%) resultó mayor que la de varones (42%). En estudios previos no se han encontrado diferencias por género (Macbeth & López Alonso, 2008). La edad de los participantes registró una media de 22,37 años (DE=4,155). Los participantes se asignaron al grupo control (n=72) o al grupo experimental (n=80) de manera aleatoria. En todos los casos la participación fue libre, voluntaria y con consentimiento escrito.

Instrumento

Se administró a todos los participantes el TCR (López Alonso, 1988, 1996) para medir su éxito objetivo en tareas de razonamiento lógico. Esta prueba posee adecuadas propiedades psicométricas y ha sido aplicada en diversos estudios sobre razonamiento lógico y representaciones sociales (López Alonso, 2000). La consigna del TCR solicita definir cuatro términos y luego completar 64 ítems

que relacionan esos términos en proposiciones compuestas. Las posibles relaciones entre las que eligen los participantes son tres: necesidad, imposibilidad y contingencia (López Alonso, 1991b). Para evaluar el comportamiento de la calibración en diferentes formas proposicionales, se agruparon los 64 ítems en cuatro subescalas, como se indicó en la Tabla 1. La confiabilidad de estas subescalas resultó psicométricamente adecuada, con un coeficiente α de Cronbach superior a 0,78 en cada una. Se agregó una consigna de estimación subjetiva de éxito luego de cada subescala. Si bien el TCR en su forma original mide errores lógicos, por lo cual el máximo desempeño obtiene puntaje cero, se empleó en el presente estudio una puntuación alternativa. Se midió la cantidad de aciertos en los 64 ítems. Esta estrategia es admisible, complementaria y coherente con la versión original del TCR (López Alonso, 1988).

Procedimiento

Al grupo control se le administraron las cuatro subescalas del TCR de la siguiente manera. En un protocolo de papel y lápiz se incluyó la consigna original del TCR que solicita completar proposiciones compuestas que establecen relaciones modales entre cuatro términos definidos previamente por la persona evaluada. Para cada ítem o proposición compuesta la persona evaluada debió elegir entre las tres posibles relaciones antes indicadas, según la forma lógica que implícitamente generan las definiciones dadas. Se establecieron cortes cada 16 ítems, en correspondencia con las cuatro subescalas descritas en la Tabla 1. Luego de cada corte los participantes debían emitir un juicio de éxito subjetivo restringido a la última subescala completada. Este juicio consistía en indicar cuántas de las respuestas dadas considera la persona evaluada que han sido correctas sobre un total de 16.

Al grupo experimental se le aplicaron dos fases para lograr la manipulación, una de estudio y otra de prueba. El procedimiento en fase de prueba fue idéntico al del grupo control, sólo que se incluyeron dos subescalas (I y IV) y no las cuatro administradas al grupo control. En fase de estudio se realizó un entrenamiento con las otras dos subescalas. Para ello se solicitó que completaran sucesivamente las subescalas II y III, incluyendo una estimación subjetiva de éxito al finalizar cada una. Luego de completar estas dos subescalas, el experimentador señaló las respuestas correctas dadas en cada ítem por los participantes de estudios previos que definieron los términos de la manera más frecuente (López Alonso, 1991a). Esta modalidad típica de respuesta adopta para la definición de los términos iniciales el criterio del lugar de nacimiento. Se responde, por ejemplo, 'Una persona inglesa es aquella nacida en Inglaterra'. Esta estrategia de respuesta ha sido empleada por la totalidad de los participantes del presente estudio. La retroalimentación dada por el experimentador se reforzó indicando a los participantes que atiendan a las posibles discrepancias entre el rendimiento alcanzado y su estimación subjetiva de éxito. Se advirtió que más tarde, dentro de la misma sesión experimental, se presentarían nuevas tareas similares a la tarea entrenada.

Para operacionalizar la variable O se evaluó la cantidad de respuestas correctas obtenidas en cada subescala y para medir la variable E se evaluó la estimación personal del rendimiento alcanzado, también por subescala.

Hipótesis

Se pronostica la ocurrencia espontánea de sesgos de la calibración en el grupo control. La hipótesis experimental H1 afirma la discrepancia entre el éxito subjetivo E y el éxito objetivo O en algunas de las cuatro subescalas del TCR, en ausencia de manipulación. Formalmente,

$H1: \exists k, E_k \neq O_k, k = \{1, 2, 3, 4\}$. De manera más específica, se pronostica el sesgo de subconfianza debido a la facilidad relativa de las tareas que se ha observado en aplicaciones previas del TCR (López Alonso, 1991a). En experimentos previos de calibración se ha encontrado frecuentemente subconfianza ante tareas relativamente fáciles (Macbeth & Morán, 2009; Macbeth & Razumiejczyk, 2008).

La segunda hipótesis experimental H2 pronostica que la forma de las proposiciones indicada en la Tabla 1 resulta relevante para la ocurrencia del sesgo de subconfianza en el grupo control. En particular, se afirma que las formas que relacionan proposiciones *monótonas* o del mismo signo son

más propensas a generar el sesgo de subconfianza que las proposiciones *variantes* o de signo opuesto. Las hipótesis específicas afirman que las subescalas I y IV generan el sesgo de subconfianza en el grupo control por ser monótonas. Formalmente, para la subescala I se afirma que $H2a : E_I < O_I$ y para la escala IV se pronostica que $H2b : E_{IV} < O_{IV}$. La conjetura que sostiene estas hipótesis sugiere que la monotonía de las proposiciones compuestas genera una *ilusión de simplicidad* que reduce la intensidad del monitoreo metacognitivo durante la ejecución de la tarea. Al presentarse esta ilusión, la calibración obtiene información de menor calidad, lo cual la hace propensa a incurrir en sesgos (Santamaría & Johnson-Laird, 1998).

La tercera hipótesis experimental H3 afirma que el sesgo de subconfianza hallado en el grupo control se disuelve en el grupo experimental mediante un entrenamiento metacognitivo de dominio específico. Esta estrategia de eliminación de sesgos de la calibración ha resultado eficiente en estudios previos (Macbeth & Cortada de Kohan, 2008; Macbeth & Razumiejczyk, 2008; Macbeth & Morán, 2009). Se pronostica que la exposición de los participantes a un entrenamiento de calibración con devolución de resultados mediante las escalas II y III en fase de estudio promueve la posterior disolución del sesgo al aplicar las escalas I y IV en fase de prueba. Para las subescalas I y IV se afirma entonces la ausencia de sesgos luego del entrenamiento, es decir, $H3a : E_I \approx O_I$ y $H3b : E_{IV} \approx O_{IV}$.

Resultados

Las variables medidas resultaron incompatibles con los presupuestos de normalidad e igualdad de varianzas por las pruebas de Kolmogorov-Smirnov y Levene, respectivamente. Esto se debe a que el rendimiento en la TCR tiende a generar puntajes altos y próximos al extremo superior de las subescalas. De esta manera, la variabilidad por encima de la media presenta menor dispersión que la variabilidad por debajo de la media, generando asimetrías. Por ello se aplicaron en todos los casos pruebas no paramétricas que utilizan la mediana para la estimación de los estadísticos.

Las tres hipótesis experimentales propuestas resultaron compatibles con la evidencia obtenida. El grupo control generó el sesgo de subconfianza en dos subescalas, en coherencia con la H1. La tabla 2 presenta los estadísticos descriptivos y las pruebas de hipótesis realizadas.

Tabla 2 - La calibración en el grupo control

Estadísticos	Subescala TCR			
	I	II	III	IV
\bar{E} (DE)	13,89(3,17)	13,57(3,70)	14,04(3,15)	11,39(4,40)
\bar{O} (DE)	14,72(2,36)	13,88(3,16)	14,65(1,82)	12,60(2,92)
\bar{C} (DE)	-0,83(3,31)	-0,31(3,74)	-0,61(2,84)	-1,21(9,93)
<i>p</i> – valor	0,015*	0,360	0,297	0,021*
Calibración	subconfianza	sin sesgo	sin sesgo	subconfianza
δ Cliff	0,91	0,02	0,04	0,95

Nota: El p-valor declarado informa la significación estadística del test del signo para la diferencia entre el éxito subjetivo E y el éxito objetivo O. Si bien el test del signo emplea la mediana para la estimación del p-valor se informa aquí la media con su correspondiente desvío estándar (DE) por ser, en este caso, una medida más representativa para apreciar las diferencias. El signo * indica diferencia significativa al 0,05. La δ de Cliff indica el tamaño del efecto no paramétrico como estimador de la magnitud de posibles sesgos. Se informa aquí el valor absoluto de la δ porque en este estadístico no interesa el signo sino la medida en que se aleja de cero y se aproxima a 1, siendo 1 la máxima diferencia entre los grupos observados. A mayor δ , mayor es la magnitud del sesgo.

En coherencia con las hipótesis H2a y H2b, las subescalas que produjeron el sesgo de sobreconfianza son las monótonas. La escala I relaciona dos afirmaciones y la escala IV relaciona dos negaciones, como se indicó en la Tabla 1. Se propone explicar este resultado por una atribución espontánea y errónea de simplicidad relacionada con la invariancia de las proposiciones monótonas. Se conjetura que la asociación ecológica entre la simplicidad de una tarea y su correspondencia con un comportamiento habitualmente adaptativo produce en este caso un incremento erróneo del éxito subjetivo. Tal estrategia resulta habitualmente adaptativa, pero no en esta tarea experimental. Para profundizar el análisis de este hallazgo se propone evaluar la simplicidad de estas proposiciones de dos maneras, una empírica y otra formal. Empíricamente se encontró una diferencia significativa entre el éxito objetivo O de las cuatro subescalas mediante el test de Friedman. Un análisis *post-hoc* de a pares mediante el test del signo mostró que sólo la subescala IV resultó más difícil que cada una de las demás. El resto de las comparaciones arrojó diferencias no significativas. Esto sugiere que la ilusión de simplicidad asociada a la monotonía de las proposiciones compuestas no se corresponde con la dificultad objetiva de los razonamientos. Para evaluar esta dificultad se calculó adicionalmente el tamaño del efecto para las diferencias de a pares del éxito objetivo. Se encontró que la máxima diferencia corresponde a las subescalas I y IV (δ Cliff=0,54). Este resultado sugiere que la subescala IV resultó ser la más difícil y la I, en cambio, la más fácil. Esta evaluación empírica es coherente con una evaluación matemática de la forma de las subescalas. Si se asume que la negación, entendida como operador lógico, es una función sentencial que opera sobre una proposición previa (Tarski, 1941), se concluye que la conjunción de dos negaciones consume mayor procesamiento algorítmico que la proposición original. De esta manera, la máxima dificultad matemática corresponde a la subescala IV porque incluye dos negaciones y la mínima a la escala I porque incluye dos afirmaciones. Las subescalas II y III presentan una dificultad intermedia porque se componen de una afirmación y una negación. Estos resultados, tomados en conjunto, sugieren que el comportamiento de la calibración en tareas de razonamiento lógico no depende de la dificultad empírica ni computacional de las inferencias, sino de la activación de un *heurístico de monotonía* que genera una *ilusión de simplicidad* responsable de la reducción de intensidad del monitoreo metacognitivo. Debido a que la estimación subjetiva de éxito procesa información metacognitiva (Macbeth & Cortada de Kohan, 2008), resulta razonable que la calibración se distorsione por una reducción del monitoreo.

La hipótesis H3 resultó compatible con la evidencia experimental. La disolución de sesgos de la calibración lograda mediante entrenamiento metacognitivo de dominio específico replica hallazgos previos (Macbeth & Morán, 2009; Macbeth & Razumiejczyk, 2008). La exposición de los participantes del grupo experimental a un entrenamiento metacognitivo con devolución realizado mediante las subescalas II y III en fase de estudio logró eliminar el sesgo de subconfianza en las subescalas I y IV durante la fase de prueba. La Tabla 3 resume los estadísticos del grupo experimental luego del entrenamiento.

Tabla 3 - Disolución del sesgo de subconfianza en el grupo experimental

Estadísticos	Subescala TCR	
	I	IV
\bar{E} (DE)	13,15(3,576)	12,78(3,742)
\bar{O} (DE)	13,60(3,279)	13,63(2,558)
\bar{C} (DE)	-0,45(3,138)	-0,85(3,135)
<i>p</i> – valor	0,522	0,281
Calibración	sin sesgo	sin sesgo
$ \delta \text{ Cliff} $	0,03	0,04

Nota: se aplican las mismas especificaciones que en la Tabla 2.

La magnitud del sesgo se redujo hasta valores muy próximos a cero según el valor absoluto de la δ de Cliff, lo cual sugiere que el éxito subjetivo y el éxito objetivo no difieren por efecto del entrenamiento metacognitivo en calibración, en coherencia con el p-valor no significativo arrojado por el test del signo.

Discusión

Si bien los modelos normativos del pensamiento humano se han extendido ampliamente en la investigación cognitiva sobre el razonamiento inductivo, el razonamiento deductivo, la resolución de problemas y, sobre todo, en la toma de decisiones en situación de incertidumbre, durante los últimos años se ha vuelto necesario revisar los patrones contra los que se compara el procesamiento humano y espontáneo de la información (Dhmi, Hertwig & Hoffrage, 2004). Pareciera que la confrontación del razonamiento deductivo espontáneo con modelos formales de la lógica proposicional no ha logrado generar teorías psicológicas con poder explicativo o tecnológico de relevancia (Santamaría & Johnson-Laird, 1998). Es por ello que durante las últimas dos o tres décadas se han comenzado a estudiar no sólo las violaciones normativas a las reglas de la lógica, sino también posibles estrategias concretas que se disparan ante tareas experimentales de razonamiento. Se ha encontrado que, bajo ciertas condiciones, estas estrategias espontáneas son eficientes y adaptativas (Kahneman & Klein, 2009).

Los hallazgos del presente estudio sugieren que la calibración en tareas de razonamiento deductivo puede explicarse, en parte, por la activación de un *heurístico de monotonía* que produce una *ilusión de simplicidad*. La evidencia experimental y el análisis empírico y computacional de la dificultad de las tareas respaldan esta conclusión. Se observó que el éxito subjetivo presenta un sesgo de subestimación cuando las tareas lógicas incluyen proposiciones compuestas del mismo signo. Cuando el signo es variante, entonces no se presentan sesgos. Se sugirió que este fenómeno ocurre porque la ilusión de simplicidad provoca una disminución del monitoreo que, a su vez, reduce la calidad de la información que alimenta los procesos metacognitivos de la calibración. En el mismo sentido, se encontró que un entrenamiento metacognitivo de dominio específico logra disolver el sesgo de subconfianza en tareas de razonamiento deductivo. Esta evidencia replica hallazgos obtenidos previamente en otros dominios y resulta compatible con el enfoque de la monotonía y la simplicidad por tratarse en todos estos casos de un complejo procesamiento metacognitivo (Fernández-Duque & Black, 2007; Flavell, 1979; Nelson, 1996).

Se requieren nuevos experimentos para poner a prueba estos hallazgos. En particular, se necesitan nuevas mediciones de la calibración en tareas alternativas que utilicen proposiciones compuestas monótonas y variantes. Resulta asimismo relevante estudiar el comportamiento de la calibración en idénticas tareas, pero en condición de abstracción, es decir, mediante expresiones formales sin contenido (Evans, 1984). La comparación del razonamiento abstracto con proposiciones declarativas y fácticas permitirá explorar si el conocimiento del mundo asociado a las proposiciones de la tarea resulta o no crítico para la calibración, como se conjetura desde el enfoque ecológico (Dhmi et al., 2004; Gigerenzer et al., 1991). Se espera en futuras investigaciones continuar con el estudio del problema específico de las distorsiones del éxito subjetivo en tareas de razonamiento abstracto. Se espera asimismo que los hallazgos del presente experimento colaboren con el acercamiento reactivado recientemente por Kahneman & Klein (2009) entre los modelos normativos y los modelos ecológicos del pensamiento humano, cuyo antagonismo ha generado más discusiones inconducentes que avances en la explicación del funcionamiento de la mente humana, sus estrategias y sus distorsiones.

Referencias

- Asensio, M., Cordero, J.M., García Madruga, J.A. & Recio, J. (1990). "Ningún Iroqués era Mohicano": La influencia del contenido en las tareas de razonamiento lógico. *Estudios de Psicología*, 43, 35-60.
- Brenner, L. (2003). A random support model of the calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes*, 90, 87-110.
- Brenner, L., Koehler, D.J., Liberman, V. & Tversky, A. (1996). Overconfidence and probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes*, 65(3), 212-219.
- Camerer, C.F. & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.
- Dasí, C. & Algarabel, S. (2003). Influencia del entrenamiento sobre el razonamiento deductivo: importancia del contenido y transferencia de dominios. *Psicothema*, 15(3), 440-445.
- Dhmi, M.K., Hertwig, R. & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130(6), 959-988.
- Evans, J.B.T. (1984). Heuristics and analysis processes in reasoning. *British Journal of Psychology*, 75(4), 451-468.
- Fernandez-Duque, D. & Black, S. (2007). Metacognitive judgment and denial of deficit: Evidence from frontotemporal dementia. *Judgment and Decision Making*, 2(5), 359-370.
- Fischhoff, B. (1982). Debiasing. En D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 422-444). Cambridge: Cambridge University Press.
- Flavell, J.H. (1979). Metacognition and cognitive monitoring. A new area of cognitive development inquiry. *American Psychologist*, 34(10), 906-911.
- Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, 98(4), 506-528.
- Grossman, S.I. (2007). *Elementary Linear Algebra With Applications*. Londres: Thomson Learning.
- Juslin, P., Winman, A. & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review*, 107, 384-396.
- Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise. A failure to disagree. *American Psychologist*, 64(6), 515-526.
- Kahneman, D. & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. En D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 414-421). Cambridge: Cambridge University Press.
- Klayman, J., Soll, J.B., Gonzalez-Vallejo, C. & Barlas, S. (1999). Overconfidence: It depends on how, what and whom you ask. *Organizational Behavior and Human Decision Process*, 79, 216-247.
- Koehler, D.J., Brenner, L. & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. En T. Gilovich, D. Griffin y D. Kahneman (Eds.), *Heuristics and Biases. The Psychology of Intuitive Judgment* (pp. 686-715). Cambridge: Cambridge University Press.
- Lichtenstein, S., Fischhoff, B. & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. En D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 306-334). Cambridge: Cambridge University Press.

- López Alonso, A.O. (1988). Razonamiento humano: Un test para su consistencia interna. *Signos Universitarios*, 13, 101-179.
- López Alonso, A.O. (1991a). Incidencia de la codificación de las premisas en el logro de un razonamiento coherente. *Interdisciplinaria*, 10(2), 137-167.
- López Alonso, A.O. (1991b). Razonamiento humano: Distintos niveles de insight, estructura y organización lógica subyacente (Versión cognitiva). *Revista de Psicología General y Aplicada*, 44(2), 221-230.
- López Alonso, A.O. (1996). La medición de la organización lógica del pensamiento. En L. Pasquali (Org.), *Teoría e Métodos de Medida em Ciências do Comportamento* (pp. 225-260). Brasília, Brasil: Instituto nacional de estudos e pesquisas educacionais.
- López Alonso, A.O. (2000). Los diversos significados inferenciales de las representaciones del imaginario social. *Interdisciplinaria*, 17(1), 39-59.
- Macbeth, G. (2009a). Efecto amplificador del heurístico de anclaje sobre los sesgos de la calibración. *Acta Psiquiátrica y Psicológica de América Latina*, 55(1), 3-12.
- Macbeth, G. (2009b). Distorsiones del éxito subjetivo en la resolución de problemas matemáticos. *Boletín de Psicología*, 95(1), 59-72.
- Macbeth, G. & Cortada de Kohan, N. (2008). Efecto del entrenamiento sobre la calibración subjetiva de éxito en tareas verbales. *Revista Latinoamericana de Psicología*, 40(1), 9-20.
- Macbeth, G., Cortada de Kohan, N., Razumiejczyk, E. & López Alonso, A.O. (2006). Los sesgos de sobreconfianza y subconfianza en tareas de conocimientos generales. *Acta Psiquiátrica y Psicológica de América Latina*, 52(4), 221-226.
- Macbeth, G. & Fernández, H. (2008). Moderadores metacognitivos del sesgo de subconfianza. *Informes Psicológicos*, 10(11), 189-206.
- Macbeth, G., Ledesma, R., Razumiejczyk, E., Cortada de Kohan, N. & López Alonso, A.O. (2008). La medición de los sesgos de la calibración mediante modelos discretos y continuos. *Investigaciones en Psicología*, 13(1), 117-134.
- Macbeth, G. & López Alonso, A.O. (2008). Aportes de enfoque ecológico a los estudios sobre calibración. *Acta Psiquiátrica y Psicológica de América Latina*, 54(1), 55-61.
- Macbeth, G. & Morán, V. (2009). El sesgo de subconfianza como fenómeno de dominio específico. *Revista Latinoamericana de Psicología*, 41(1), 47-57.
- Macbeth, G. & Razumiejczyk, E. (2008). Disolución del sesgo de subconfianza en tareas verbales. *Anales de psicología*, 24(1), 143-149.
- Macbeth, G., Razumiejczyk, E. & Cortada de Kohan, N. (2006). El sesgo de sobreconfianza en tareas verbales y matemáticas. *Investigaciones en Psicología*, 11(3), 47-58.
- Merkle, E. & Van Zandt, T. (2006). An application of the Poisson Race Model to confidence calibration. *Journal of Experimental Psychology: General*, 135(3), 391-408.
- Nelson, T.O. (1996). Consciousness and metacognition. *American Psychologist*, 51(2), 102-116.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology*, 29, 261-265.
- Razumiejczyk, E., Pereyra, C. & Macbeth, G. (2009). El juicio de sentimiento de conocimiento en la identificación de estímulos gustativos. *Boletín de Psicología*, 96(2), 66-77.
- Santamaría, C. & Johnon-Laird, P.N. (1998). Ilusiones en el razonamiento proposicional. *Ponencia presentada en la I Jornada de Psicología del Pensamiento*. Santiago de Compostela: Universidad de Santiago de Compostela.

- Simmons, J.P. & Nelson, L.D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General*, 135(3), 409-428.
- Spivak, M. (1992). *Calculus. Second edition*. New York: W.A. Benjamin.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47, 143-148.
- Tarski, A. (1941). *Introduction to Logic and to the Methodology of Deductive Sciences*. New York: Oxford University Press.
- Valiña, M.A. & de Vega, M. (1988). Un estudio experimental del razonamiento cotidiano en tareas de silogismos: una aproximación pragmática. *Cognitiva*, 1, 33-62.
- Wilson, T.D., Centerbar, D.B. & Brekke, N. (2002). Mental contamination and the debiasing problem. En T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases. The Psychology of Intuitive Judgment* (pp. 185-200). Cambridge: Cambridge University Press.