

El sesgo de subconfianza como fenómeno de dominio específico

Underconfidence bias as a specific-domain phenomenon

Recibido: noviembre de 2008.

Aprobado: enero de 2009.

Guillermo Macbeth

Universidad del Salvador, Argentina

Valeria Morán

Universidad Nacional de Cuyo, Argentina

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 1312, 1° piso, C1058AAV Buenos Aires, Argentina guillermo.macbeth@mail.salvador.edu.ar. Teléfono: +54 11 4813 3400, Fax: +54 11 4813 3404.

Resumen

Se define la calibración o confianza como la coincidencia o discrepancia entre el éxito subjetivo y el éxito objetivo que presenta una persona en la realización de una colección de tareas. Las discrepancias entre uno y otro generan distorsiones conocidas como sesgos de la calibración. Cuando el éxito subjetivo predomina sobre el éxito objetivo se observa el sesgo de sobreconfianza. En el caso contrario, se observa el sesgo de subconfianza. Se presenta un estudio experimental que intenta decidir si el dominio es relevante para la disolución del sesgo de subconfianza. Los resultados indican que el dominio es efectivamente relevante para reducir y eliminar esta distorsión del pensamiento humano. La exposición de los participantes a un modelo metacognitivo de calibración de dominio específico logró disolver el sesgo. Cuando se empleó un modelo de dominio no específico, en cambio, el sesgo no se disolvió sino que se invirtió produciendo el sesgo de sobreconfianza. Se concluye que el sesgo de subconfianza es un fenómeno metacognitivo de dominio específico.

Palabras clave: calibración, dominio general, dominio específico, sesgo de subconfianza, metacognición.

Abstract

Calibration or confidence can be defined as the empirical convergence or divergence between the subjective success and the objective success achieved by a person in a series of tasks. Discrepancies between the two generate distortions known as calibration biases. The overconfidence bias is observed when the subjective success is greater than the objective success. The underconfidence bias is defined when the inverse case is observed. This article reports an experiment that studies the relevance of the domain for debiasing the underconfidence bias. The aim of the study is to decide if calibration phenomena have specific-domain or general-domain architecture. Results show that domain specificity is relevant for the dissolution of these human thought distortions. The exposure of participants to specific domain calibration training achieved the debiasing purpose. When non-specific domain training was applied, the overconfidence bias was obtained. It is concluded that the underconfidence bias is a specific domain phenomenon.

Key words: calibration, general-domain, specific-domain, underconfidence bias, metacognition

Calibración, sobreconfianza, subconfianza

La calibración o confianza C puede definirse como la relación de convergencia o divergencia empírica entre el éxito subjetivo o estimado E y el éxito real u objetivo O que presenta una persona en la realización de una serie de tareas (Brenner, Koehler, Liberman & Tversky, 1996; Camerer & Lovallo, 1999; Gigerenzer, Hoffrage & Kleinbölting, 1991; Hausmann & Läge, 2008; Oskamp, 1965). Un estudiante universitario, por ejemplo, puede estimar que su rendimiento en un examen ha sido muy alto cuando en realidad ha sido más bien bajo, es decir, su éxito subjetivo es mayor que su éxito objetivo. Otro estudiante, por el contrario, puede considerar que no logró aprobar el examen cuando en realidad su calificación resultó muy alta, es decir, su éxito subjetivo es menor que su éxito objetivo. Finalmente, otro estudiante pudo obtener el resultado exacto que su estimación subjetiva de éxito le indicó. En el primer caso se observa el sesgo de sobreconfianza ($E > O$), mientras que en el segundo se produjo el fenómeno contrario, conocido como sesgo de subconfianza ($E < O$). En el tercer caso, en cambio, el estudiante se encuentra bien calibrado porque su confianza resultó libre de sesgos ($E = O$).

Esta variedad retrospectiva de calibración puede ser estudiada como un fenómeno complejo de pensamiento que involucra la participación de procesos cognitivos y metacognitivos o de recursividad (Fernandez-Duque & Black, 2007; Macbeth & Cortada de Kohan, 2008). Oskamp (1965) ha propuesto definir la calibración C según la Ecuación 1, cuyo valor resulta positivo para la sobreconfianza y negativo para la subconfianza. Para obtener la calibración C se calculan las discrepancias entre el éxito subjetivo E y el éxito objetivo O para una colección de n cantidad de tareas (Bersabé Morán, Martínez Arias & Tejeiro Salguero, 2003; Macbeth, Ledesma, Razumiejczyk, Cortada de Kohan & López Alonso, 2008).

$$C = \sum_{i=1}^n E_i - O_i \quad (1)$$

Se han estudiado diversas variables de la tarea, del participante y del contexto para estos fenómenos de la calibración (Brenner, 2003; Klayman, Soll, Gonzalez-Vallejo & Barlas, 1999; Koehler, Brenner & Griffin, 2002; Lichtenstein, Fischhoff & Phillips, 1982; Merkle & Van Zandt, 2006). Entre los hallazgos más relevantes para el presente estudio se encuentran: 1) el efecto difícil-fácil

(Lichtenstein *et al.*, 1982; Macbeth, Cortada de Kohan, Razumiejczyk & López Alonso, 2006), 2) la disolución de la subconfianza mediante entrenamiento metacognitivo en calibración (Macbeth & Razumiejczyk, 2008) y 3) la importancia ecológica de los modelos mentales para el logro de una calibración libre de sesgos (Gigerenzer *et al.*, 1991; Macbeth & López Alonso, 2008; Mellers, Hertwig & Kahneman, 2001).

El efecto difícil-fácil consiste en la intensificación de la sobreconfianza cuando las tareas son cada vez más difíciles y de la subconfianza cuando las tareas son cada vez más fáciles (Gilovich, Griffin & Kahneman, 2002; Juslin, Winman & Olsson, 2000; Kahneman & Tversky, 2000; Macbeth *et al.*, 2006). Este fenómeno se ha observado en tareas del dominio económico (Angner, 2006; Camerer & Lovallo, 1999), matemático (Macbeth, Cortada de Kohan & Kohan Cortada, 2007), físico (Henrion & Fischhoff, 1986), verbal (Macbeth & Cortada de Kohan, 2008; Macbeth, Razumiejczyk & Cortada de Kohan, 2006), de diagnóstico psicológico (Oskamp, 1965) y de conocimientos generales (Macbeth *et al.*, 2006), entre otros (Simmons & Nelson, 2006; Svenson, 1981) tanto dentro, como fuera del laboratorio (Christensen-Szalanski & Bushyhead, 1981; Clayson, 2005; Koehler *et al.*, 2002). El efecto difícil-fácil pronostica que tanto las tareas fáciles, como los participantes de alto rendimiento, tienden a presentar el sesgo de subconfianza. Ambos fenómenos resultan similares en tanto implican puntajes altos de éxito objetivo O en comparación con el éxito subjetivo E (Macbeth *et al.*, 2006).

La disolución del sesgo de subconfianza (Fischhoff, 1982; Kahneman & Tversky, 1982) se ha logrado en estudios recientes mediante un entrenamiento metacognitivo sencillo (Macbeth & Cortada de Kohan, 2008; Macbeth & Razumiejczyk, 2008). La calibración libre de sesgo se obtuvo mediante la exposición de los participantes en fase de estudio a un ejercicio de estimación subjetiva de éxito con devolución de los resultados. Se empleó el mismo dominio de tareas para las fases de estudio y prueba. Esta manipulación permitió corregir la distorsión de la calibración. Sin embargo, no se ha intentado aún la disolución experimental del sesgo de subconfianza mediante un entrenamiento que, en fase de estudio, emplee tareas de un dominio diferente del dominio de la fase de prueba.

Los aportes del enfoque ecológico a los estudios sobre calibración han destacado la importancia de la informa-

ción contextual para la configuración del éxito subjetivo (Dhimi, Hertwig & Hoffrage, 2004; García-Retamero & Dieckmann, 2006). Al respecto, Gigerenzer *et al.* (1991) sostienen que ante una tarea de estimación subjetiva de éxito, los participantes construyen modelos mentales que se alimentan de varias fuentes de información (Clare & Parrott, 1994; Johnson, Hashtroudi & Lindsay, 1993). El recuerdo de experiencias previas similares a la tarea experimental se combina con procesos inferenciales probabilísticos para generar juicios de éxito subjetivo (Macbeth & López Alonso, 2008). Estos experimentos ecológicos sugieren que la exposición a modelos o ejemplos de calibración en fase de estudio resulta relevante para la estimación de éxito subjetivo en fase de prueba (García-Retamero & Dhimi, 2009).

Metacognición, dominios y calibración

La calibración puede ser considerada como un fenómeno complejo (Montealegre, 2007) que involucra la participación de procesos cognitivos y metacognitivos (Fernandez-Duque & Black, 2007). La metacognición puede ser entendida, en términos generales, como un proceso mental recursivo de monitoreo y control de otros procesos cognitivos (Organista Díaz, 2005). Varios estudios han sugerido que la metacognición abarca simultáneamente una diversidad de dominios, por lo cual puede ser caracterizada como un proceso de dominio general (Fernandez-Duque & Black, 2007; Osborne, 1998; Schraw, Dunkle, Bendixen & de Backer Roedel, 1995). Desde esta perspectiva resulta coherente esperar que un entrenamiento en calibración sobre un dominio no verbal genere la reducción o disolución del sesgo de subconfianza en un dominio verbal específico.

Estos pronósticos, sin embargo, resultan incompatibles con estudios recientes sobre disolución de la calibración en relación con la experticia en tareas matemáticas y verbales (Macbeth *et al.*, 2006, 2007). Diversos experimentos realizados con expertos en estos dominios han encontrado que los sesgos de la calibración se presentan tanto en el dominio de la experticia como en el dominio de la no experticia.

La revisión conjunta de estos antecedentes recomienda estudiar de manera específica la relevancia del dominio para la disolución de los sesgos de la calibración porque algunos autores consideran que el dominio es relevante

(Fernandez-Duque & Black, 2007; Schraw *et al.*, 1995) y otros consideran que no lo es (Kelemen, Frost & Weaver, 2000). En tal sentido, la contribución particular del presente estudio radica en la comparación experimental de dos protocolos de entrenamiento en calibración que permite decidir si el dominio específico es más eficiente que el dominio no específico para corregir el sesgo de subconfianza. En estudios previos se han empleado participantes que adquirieron experticia en un dominio determinado antes del experimento, pero no se han realizado aún comparaciones de entrenamientos específicos e inespecíficos adquiridos durante el experimento, según se indica en revisiones recientes (Gilovich *et al.*, 2002; Macbeth & López Alonso, 2008).

Asimismo, el problema del dominio para la configuración de la calibración presenta interés aplicado (Armor & Taylor, 2002; Fischhoff, 1982; Macbeth & Bogaizian, 2007; Wilson, Centerbar & Brekke, 2002). Cualquier programa de disolución de estos sesgos en el ejercicio de diversas actividades y profesiones requiere un modelo teórico que explique la relación entre calibración y dominio (Brenner, 2000; Erev, Wallsten & Budescu, 1994; Kahneman, 2003; Kahneman, Slovic & Tversky, 1982; Kahneman & Tversky, 2000; Koehler *et al.*, 2002; Wallsten, Erev & Budescu, 2000).

En este contexto resulta importante estudiar si la calibración posee una arquitectura de dominio específico o si se configura como un conjunto de fenómenos de dominio general que atraviesa una multiplicidad de procesos cognitivos. La solución de este problema teórico en relación con el sesgo de subconfianza permitiría implementar intervenciones correctivas para diversas actividades (García-Retamero & Dhimi, 2009). Por ejemplo, en el dominio educativo permitiría desarrollar didácticas enriquecidas por el monitoreo metacognitivo libre de sesgos del éxito subjetivo que el alumno realiza sobre su propio aprendizaje (Clayson, 2005). En el dominio económico permitiría que los asesores financieros adecuadamente entrenados cometan menos errores y generen menos pérdidas de capitales (Angner, 2006; Camerer & Lovallo, 1999). En el dominio médico, por su parte, se haría posible la optimización sistemática de la exactitud del diagnóstico clínico mediante programas de entrenamiento en calibración desarrollados sobre la toma de decisiones médicas (Koehler *et al.*, 2002; Oskamp, 1965).

Método

Participantes

Participaron en este experimento 223 estudiantes de *Nivel Polimodal* del Sistema Educativo Argentino que abarca desde los 15 hasta los 18 años de edad y cuenta con una escolarización previa de nueve ciclos lectivos. Todos los participantes se reclutaron en la Ciudad de Mendoza, Argentina. Predominaron las mujeres ($n = 175$; 78,5%) sobre los varones ($n = 48$; 21,5%). No se consideraron en este estudio las diferencias entre géneros porque en diversos experimentos previos las discrepancias de calibración entre mujeres y varones resultaron no significativas (Camerer & Lovallo, 1999; Bersabé Morán *et al.*, 2003). La media de edad de los participantes resultó de 15,85 años ($DE = 0,89$) y la mediana de 16 años. Se establecieron tres grupos formados por dos grupos experimentales etiquetados en lo que sigue como EXP1 ($n = 73$) y EXP2 ($n = 72$) y uno control ($n = 78$). Se mantuvieron proporciones similares de varones y mujeres en los tres grupos. La asignación de los participantes a cada grupo fue aleatoria. La participación en el experimento fue libre, voluntaria, con conocimiento y consentimiento escrito de cada participante.

Instrumentos

Se administraron dos instrumentos con adecuadas propiedades psicométricas, una prueba de aptitud verbal (BAIRES) de Cortada de Kohan (2003) y una prueba de conocimientos generales (Macbeth *et al.*, 2006).

La prueba BAIRES mide el desempeño de jóvenes y adultos en tareas de sinónimos y definiciones en lengua española. Esta prueba se administró a los tres grupos de participantes. BAIRES es un instrumento de papel y lápiz construido mediante las técnicas de la Teoría de Respuesta al Ítem. La confiabilidad de los puntajes obtenidos en esta prueba se mantuvo dentro de lo esperado (α de Cronbach = 0,637). Las tareas específicas del instrumento presentan el formato de respuesta múltiple, con cuatro opciones de respuesta para cada pregunta, de las cuales siempre una y sólo una es la correcta. La versión extensa del BAIRES que se aplicó en fase de prueba a los tres grupos de participantes cuenta con 34 ítems. Los 17 primeros ítems presentan tareas de definiciones y los 17 restantes, de sinónimos. Para las definiciones se presenta un

sustantivo acompañado por cuatro definiciones posibles. Para los sinónimos se presenta un sustantivo y cuatro sinónimos posibles. Se solicita a los participantes que elijan la opción que, en cada caso, consideran más acertada.

Una versión abreviada del mismo instrumento (Cortada de Kohan, 2003; Macbeth & Cortada de Kohan, 2008) se aplicó en fase de estudio, a modo de entrenamiento de dominio específico, a los participantes del grupo EXP1. La versión abreviada del BAIRES está compuesta por 16 ítems, de los cuales 8 son tareas de definiciones y 8 son tareas de sinónimos. Se empleó esta versión abreviada en fase de estudio para evitar una sobrecarga cognitiva de la memoria de trabajo en fase de prueba, en coherencia con estudios previos (Macbeth & Cortada de Kohan, 2008; Macbeth & Razumiejczyk, 2008). Las propiedades psicométricas de la versión abreviada son homogéneas con las propiedades de la versión extensa. Las versiones abreviada y extensa no comparten ningún ítem.

También se empleó una prueba de conocimientos generales (Macbeth *et al.*, 2006) que posee características de administración y propiedades psicométricas similares al BAIRES. Esta prueba, que se aplicó sólo al grupo EXP2 en fase de estudio, cuenta con 16 preguntas de conocimientos generales que abarcan temas de historia, geografía, ciencia y arte. Se incluyeron cuatro preguntas por tema. Cada pregunta presenta cuatro opciones de respuesta, de las cuales una y sólo una es la correcta. La confiabilidad de este cuestionario resultó adecuada (α de Cronbach = 0,68).

Ambas pruebas son válidas, confiables y han sido empleadas en estudios previos sobre calibración (Macbeth *et al.*, 2006; Macbeth & Cortada de Kohan, 2008).

Procedimiento

Al grupo control ($n = 78$) se le administró sólo la versión extensa del BAIRES en fase de prueba. Al grupo EXP1 ($n = 73$) se le administró, en fase de estudio, el BAIRES abreviado con las respuestas dadas por una persona ficticia, a modo de entrenamiento de dominio específico, y la versión extensa del mismo instrumento en fase de prueba. Al grupo EXP2 ($n = 72$), en cambio, se le administró en fase de estudio el cuestionario de conocimientos generales con las respuestas dadas por una persona ficticia, a

modo de entrenamiento de dominio no específico y, en fase de prueba, la versión extensa del BAIRES.

Los ejemplos o modelos presentados en ambos grupos experimentales en fase de estudio incluían las respuestas dadas por una persona ficticia. La tarea de los participantes ante este ejemplo consistía en estimar cuántas de las 16 respuestas dadas por esta persona ficticia son correctas. En los dos ejemplos, específico y no específico, todas las respuestas dadas eran correctas. El objetivo de esta consigna fue que los participantes respondieran espontáneamente una cifra menor que 16. El grupo EXP1 obtuvo en esta variable una media de 13,19 ($DE = 1,94$) y el grupo EXP2 obtuvo una media de 12,87 ($DE = 2,04$). La diferencia entre ambos resultó no significativa ($t = 0,958$; $p = 0,339$; d de Cohen = 0,16). Luego, al final de la fase de estudio, se pretendió generar en los participantes la sensación de haber incurrido en una subestimación al revelar el experimentador la cantidad verdadera de respuestas correctas del ejemplo (16/16) (Clare & Parrott, 1994; Johnson *et al.*, 1993). Se buscó que esta información ingresada en fase de estudio, es decir, la sensación de haber subestimado, generara más tarde, en fase de prueba, un incremento experimental del éxito subjetivo luego de completar la versión extensa del BAIRES. Esta manipulación es coherente con los aportes del enfoque ecológico a los estudios sobre calibración que destacan la relevancia de la información contextual para el ajuste adaptativo del éxito subjetivo. No se implementaron tareas distractoras entre la fase de estudio y la fase de prueba.

Se administró el instrumento en forma grupal. Se realizaron en total 6 sesiones para la recolección de datos ($30 < n < 40$) en condiciones experimentales homogéneas. Los participantes de los tres grupos completaron la versión extensa del BAIRES en aproximadamente 20 minutos en fase de prueba. El BAIRES abreviado y el cuestionario de conocimientos generales aplicados a los grupos experimentales en fase de estudio no consumieron más de 15 minutos en ninguna sesión.

Hipótesis

La hipótesis H_1 afirma que el grupo control presenta, en fase de prueba, el sesgo de subconfianza si su rendimiento es relativamente alto o el sesgo de sobreconfianza si su rendimiento es relativamente bajo. Formalmente,

$$H_1 : \begin{cases} \bar{O}_{alto} \rightarrow \bar{E}_{CONTROL} < \bar{O}_{CONTROL} \\ \bar{O}_{bajo} \rightarrow \bar{E}_{CONTROL} > \bar{O}_{CONTROL} \end{cases}$$

Este pronóstico se justifica por el efecto difícil-fácil informado en estudios previos (Gigerenzer *et al.*, 1991; Macbeth *et al.*, 2006; Oskamp, 1965). En este experimento se esperaba observar, específicamente, el sesgo de subconfianza porque el BAIRES había producido rendimientos altos en experimentos previos sobre calibración (Macbeth & Cortada de Kohan, 2008; Macbeth & Razumiejczyk, 2008).

La hipótesis H_2 afirma que el grupo EXP1 no presenta sesgos de la calibración en fase de prueba. Este grupo experimental se definió por su exposición, en fase de estudio, a un ejemplo de *dominio específico* que incrementa el éxito subjetivo E . Formalmente, $H_2 : \bar{E}_{EXP1} \approx \bar{O}_{EXP1}$. Estudios previos han informado que el sesgo de subconfianza se disuelve mediante entrenamiento metacognitivo en tareas de calibración del mismo dominio y que esta disolución se debe al entrenamiento en calibración antes que a la adquisición de experticia en la tarea (Macbeth & Cortada de Kohan, 2008). El entrenamiento del grupo EXP1 fue sin embargo, novedoso en tanto utilizaba un ejemplo de respuestas dadas por una persona ficticia en una tarea de dominio específico. Esta exposición permitió manipular el éxito subjetivo E de manera focalizada.

La hipótesis H_3 sostiene que el grupo EXP2 presenta sesgos de la calibración en fase de prueba. Este grupo experimental se definió por su exposición, en fase de estudio, a un ejemplo de *dominio no específico* que incrementa el éxito subjetivo E en caso de observarse el sesgo de subconfianza en el grupo control, según se indicó en H_1 . Se esperaba que el entrenamiento metacognitivo en calibración implementado mediante un ejemplo de dominio no específico generara el sesgo de sobreconfianza en fase de prueba por un incremento errático y exagerado del éxito subjetivo E . Se pronosticó que esto se producirá por la exposición de los participantes, en fase de estudio, a un ejemplo de calibración sobre conocimientos generales cuyas estimaciones no pueden transferirse directamente a la calibración sobre tareas verbales. Formalmente, $H_3 : \bar{E}_{EXP2} > \bar{O}_{EXP2}$

La ocurrencia de la subconfianza en el grupo control es un presupuesto de las hipótesis H_2 y H_3 . Si el grupo control, por el contrario, mostrase el sesgo de sobreconfianza, definido como una sobreestimación del éxito subjetivo, las manipulaciones experimentales de los grupos EXP1 y EXP2 propondrían un decremento antes que un incremento de la variable E y la hipótesis H_3 afirmaría la desigualdad opuesta, es decir, $H_1 : \bar{E}_{CONTROL} > \bar{O}_{CONTROL} \rightarrow H_3 : \bar{E}_{EXP2} < \bar{O}_{EXP2}$

La hipótesis H_4 es un complemento de las hipótesis H_1 , H_2 y H_3 y afirma que la calibración C presenta diferencias significativas entre los tres grupos en fase de prueba. Formalmente, $H_4 : \bar{C}_{EXP1} \neq \bar{C}_{EXP2} \neq \bar{C}_{CONTROL}$

Resultados

Para evaluar la eficacia de las manipulaciones experimentales se realizó un análisis preliminar del éxito objetivo y del éxito subjetivo en los tres grupos. Se controló que el

éxito objetivo O no presentara una diferencia significativa entre los tres grupos, pero que el éxito subjetivo E sí la presentara porque el experimento se propuso manipular específicamente la segunda variable y no la primera. Se procuró que las variaciones de la calibración C se explicaran por las manipulaciones experimentales del éxito subjetivo E provocadas por la presentación de ejemplos de dominio específico (EXP1) y no específico (EXP2) y no por variaciones no controladas del éxito objetivo O .

Con respecto al éxito objetivo O , los puntajes obtenidos en el BAIREs en fase de prueba por los tres grupos no presentaron diferencias significativas ($F = 1,902; gl = 2; p = 0,152; \eta_p^2 = 0,017$). Con respecto al éxito subjetivo E , en fase de prueba, las diferencias entre los grupos resultaron significativas ($F = 8,207; gl = 2; p < 0,001; \eta_p^2 = 0,069$). El análisis conjunto de estos resultados, cuyo resumen se presenta en la Tabla 1, indica que la manipulación logró generar variaciones experimentales sólo en el éxito subjetivo E sin afectar simultáneamente al éxito objetivo O , tal como lo exige el objetivo del experimento.

Tabla 1
Estadísticos descriptivos del éxito subjetivo y el éxito objetivo en los tres grupos

	Grupo		
	Experimental 1	Experimental 2	Control
Éxito subjetivo $\bar{E}(DE)$	18,29 (5,179)	18,58 (5,336)	15,59 (4,545)
Éxito objetivo $\bar{O}(DE)$	18,51 (3,271)	17,40 (3,637)	18,19 (3,615)

La hipótesis H_1 resultó coherente con la evidencia experimental. El grupo control presentó el sesgo de subconfianza en fase de prueba. El éxito subjetivo ($\bar{E} = 15,59; DE = 4,545$) resultó significativamente menor que el éxito objetivo ($\bar{O} = 18,19; DE = 3,615$) por la prueba t para muestras apareadas ($t = -5,08; p < 0,01; d$ de Cohen = 0,63). Este resultado se explica por el rendimiento relativamente alto de estos estudiantes de nivel medio si se los compara con una media de 18,85 ($DE = 4,178$) respuestas correctas alcanzada por estudiantes universitarios avanzados que participaron en otro experimento similar que aplicó el BAIREs para medir el éxito objetivo (Macbeth *et al.*, 2006).

La hipótesis H_2 resultó coherente con la evidencia experimental. El grupo EXP1 no presentó sesgos de la

calibración en fase de prueba. El grupo EXP1 fue expuesto a un ejemplo de *dominio específico* que incrementa el éxito subjetivo E . El éxito subjetivo ($\bar{E}_{EXP1} = 18,29; DE = 5,179$) y el éxito objetivo ($\bar{O}_{EXP1} = 18,51; DE = 3,271$) no discreparon significativamente ($t = -0,355; p = 0,724; d$ de Cohen = 0,05), a diferencia del grupo control que presentó el sesgo de subconfianza, según H_1 .

La hipótesis H_3 resultó coherente con la evidencia empírica. El grupo EXP2 presentó el sesgo de sobreconfianza en fase de prueba. Este grupo de participantes fue expuesto a un ejemplo de *dominio no específico* que incrementa el éxito subjetivo E , como se indica más arriba en la sección *Procedimiento*. El éxito subjetivo ($\bar{E}_{EXP2} = 18,58; DE = 5,336$) resultó significativamente mayor ($t = 2,029; p < 0,05; d$ de Cohen = 0,26) que el éxito

objetivo ($\bar{O}_{EXP1} = 17,40$; $DE = 3,637$). El sesgo de subconfianza observado en el grupo control no se observó en el grupo EXP2, pero tampoco se generó una buena calibración como en el grupo EXP1. El incremento experimental del éxito subjetivo E produjo la distorsión opuesta, es decir, el sesgo de sobreconfianza. Esto puede deberse a que la exposición a un ejemplo de dominio no específico en fase de estudio provocó, en el grupo EXP2, una *ilusión de ajuste de la calibración* en fase de prueba. El ajuste mediante un ejemplo de dominio no específico resultó ilusorio porque la calibración funciona,

según lo sugiere la evidencia, como un fenómeno de dominio específico.

La hipótesis H_4 resultó coherente con la evidencia experimental. La calibración C presentó, en fase de prueba, una diferencia significativa ($F = 11,424$; $gl = 2$; $p < 0,001$; $\eta_p^2 = 0,094$) entre los tres grupos por las manipulaciones realizadas en fase de estudio sobre el éxito subjetivo E . La Tabla 2 presenta los valores de la calibración C en cada grupo y la Figura 1 permite observar los resultados del ANOVA.

Tabla 2
Fenómenos de la calibración observados en los tres grupos

	Grupo		
	Experimental 1	Experimental 2	Control
Calibración \bar{C} (DE)	-0,22 (5,276)	1,18 (4,937)	-2,60 (4,525)
Fenómeno	buena calibración $E \approx 0$	sesgo de sobreconfianza $E > 0$	sesgo de subconfianza $E < 0$

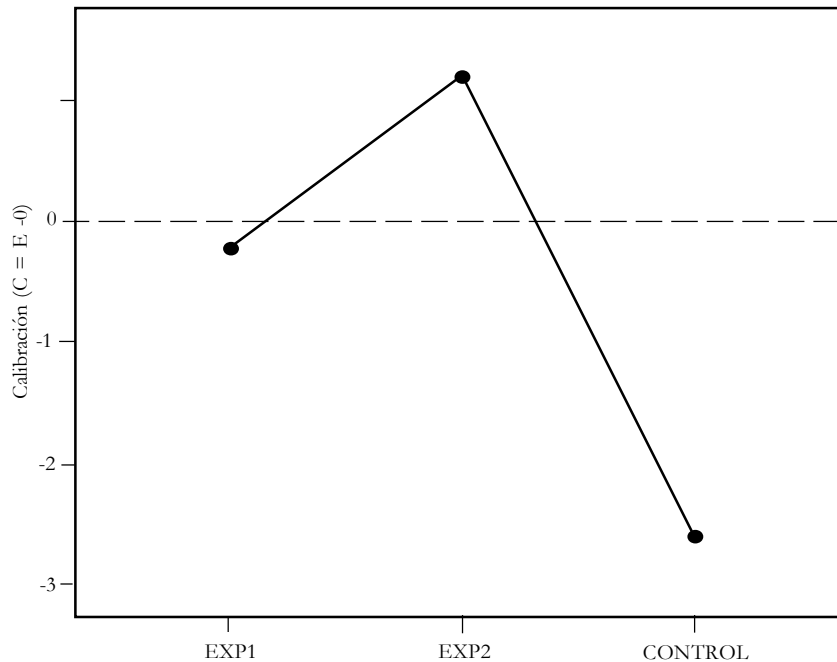


Figura 1. Aproximación de la calibración a la ausencia de sesgos ($C = \text{cero}$) en los tres grupos

Todas las variables computadas resultaron normales y homocedásticas por las pruebas de Kolmogorov-Smirnov y Levene, respectivamente.

Estos resultados sugieren, en conjunto, que el dominio de la tarea es relevante para la disolución del sesgo de subconfianza.

Discusión

Se discute, en general, si el dominio es relevante para la configuración de los fenómenos de la calibración. En el mismo sentido y de manera más específica se estudia si el dominio es crítico para la disolución del sesgo de subconfianza. Este problema es relevante a nivel teórico porque algunos modelos sugieren que la calibración como fenómeno metacognitivo posee una arquitectura de dominio general (Fernandez-Duque & Black, 2007; Osborne, 1998; Schraw *et al.*, 1995) y otros, en cambio, sostienen que la estimación subjetiva de éxito se comporta de manera diferente en diversos dominios (Kelemen *et al.*, 2000). A la vez, el problema de la calibración y los dominios presenta interés tecnológico a nivel aplicado (García-Retamero & Dhami, 2009). En varios estudios se ha señalado la posibilidad de desarrollar programas de entrenamiento cognitivo para la eliminación de las distorsiones de la calibración (Kahneman & Tversky, 1982; 2000; Koehler *et al.*, 2002). Así, se han sugerido aplicaciones en ámbitos educativos, económicos, judiciales y médicos, entre otros (Macbeth & López Alonso, 2008).

Se ha encontrado en estudios previos que la calibración se libera de sesgos mediante un entrenamiento experimental de dominio específico (Macbeth & Cortada de Kohan, 2008; Macbeth & Razumiejczyk, 2008). Sin embargo, no se registran en el estado del arte experimentos que comparen entrenamientos en calibración de dominio específico con entrenamientos de dominio general. Se propuso en el presente estudio replicar hallazgos previos y compararlos de manera novedosa con un entrenamiento de dominio no específico.

La evidencia experimental indica que el sesgo de subconfianza se disuelve mediante la exposición de los participantes a un ejemplo de calibración sobre tareas de dominio específico, en coherencia con estudios previos. Cuando se presentó un ejemplo de dominio no específico, en cambio, la calibración se comportó de manera

errática produciendo el sesgo de sobreconfianza. Se propone explicar este fenómeno como una *ilusión de ajuste de la calibración*.

Se concluye que el sesgo de subconfianza es un fenómeno de dominio específico. Esta contribución promueve el desarrollo de intervenciones correctivas específicas en ámbitos aplicados. Los resultados obtenidos permiten explicar por qué fracasaron numerosos intentos previos de corrección de sesgos de la calibración. En diversos estudios se ha intentado eliminar sin éxito la distorsión mediante la advertencia directa a los participantes (Oskamp, 1965), la recompensa con incentivos económicos (Angner, 2006; Camerer & Lovallo, 1999), la exposición a información ecológica relevante (Gigerenzer *et al.*, 1991; Macbeth & López Alonso, 2008), la adquisición de experticia sin monitoreo metacognitivo de la calibración (Christensen-Szalanski & Bushyhead, 1981; Macbeth & Cortada de Kohan, 2008) y la motivación (Clayson, 2005), entre otros (Fischhoff, 1982).

Entre las limitaciones de este estudio se destaca el empleo de un único entrenamiento de dominio no específico. Se recomienda estudiar en futuras investigaciones el comportamiento de la calibración ante diversos protocolos de entrenamiento de dominio general. Otra limitación importante consiste en la comparación de los entrenamientos correctivos con un grupo control que presenta el sesgo de subconfianza. Se recomienda replicar este experimento sobre la base de un grupo control con sesgo de sobreconfianza.

Los hallazgos del presente estudio son coherentes con aportes previos (Klayman *et al.*, 1999) y recomiendan investigar su generalización a otras tareas y contextos. Se propone explorar en estudios ulteriores si la calibración es, en general y no sólo para el sesgo de subconfianza, un fenómeno metacognitivo de dominio específico. Esta generalización implica estudiar el comportamiento de los sesgos de sobreconfianza y subconfianza, así como su ausencia en relación con el dominio para diferentes tareas, participantes y contextos.

Referencias

- Angner, E. (2006). Economists as Experts: Overconfidence in theory and practice. *Journal of Economic Methodology*, 13(1), 1-24.

- Armor, D.A. & Taylor, S.E. (2002). When predictions fail: The dilemma of unrealistic optimism. En T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases. The Psychology of Intuitive Judgment* (pp. 334-347). Cambridge: Cambridge University Press.
- Bersabé Morán, R., Martínez Arias, R. & Tejeiro Salguero, R. (2003). Risk-takers: Do they know how much of a risk they are taking? *Psychology in Spain*, 7(1), 3-9.
- Brenner, L. (2000). Should observed overconfidence be dismissed as a statistical artifact? Critique of Erev, Wallsten, and Budescu (1994). *Psychological Review*, 107(4), 943-946.
- 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 y 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.
- Christensen-Szalanski, J.J. & Bushyhead, J.B. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 928-935.
- Clayson, D.E. (2005). Performance overconfidence: Metacognitive effects or misplaced student expectations? *Journal of Marketing Education*, 27(2), 122-129.
- Clore, G.L. & Parrott, W.G. (1994). Cognitive feelings and metacognitive judgements. *European Journal of Social Psychology*, 24, 101-115.
- Cortada de Kohan, N. (2003). *BAIRES. Test de Aptitud Verbal*. Madrid: TEA.
- 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.
- Erev, I., Wallsten, T.S. & Budescu, D.V. (1994). Simultaneous over- and Underconfidence: The role of error in judgment processes. *Psychological Review*, 101(3), 519-527.
- 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.
- García-Retamero, R. & Dhmi, M. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review*, 16(1), 163-169.
- García-Retamero, R. & Dieckmann, A. (2006). Una visión crítica de los heurísticos rápidos y frugales. *Revista Latinoamericana de Psicología*, 38(3), 509-522.
- Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, 98(4), 506-528.
- Gilovich, T., Griffin, D. & Kahneman, D. (Eds.) (2002). *Heuristics and Biases. The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.
- Hausmann, D. & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision Making*, 3(3), 229-243.
- Henrion, M. & Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics*, 54, 791-797.
- Johnson, M.K., Hashtroudi, S. & Lindsay, S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3-28.
- 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. (2003). A perspective on judgment and choice. Mapping bounded rationality. *American Psychologist*, 58(9), 697-720.
- Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

- 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.
- Kahneman, D. & Tversky, A. (Eds.). (2000). *Choices, Values, and Frames*. New York: Cambridge University Press.
- Kelemen, W.L., Frost, P.J. & Weaver, C.A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory and cognition*, 28(1), 92-107.
- 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 & 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.
- Macbeth, G. & Bogiaizian, D. (2007). La estimación subjetiva de éxito en los trastornos de ansiedad. *Revista Argentina de Clínica Psicológica*, 16(2), 143-150.
- 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. & Kohan Cortada, A. (2007). Efecto de la experticia matemática en el sesgo de sobreconfianza. *Perspectivas en Psicología*, 4(1), 13-18.
- 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., 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 del enfoque ecológico a los estudios sobre calibración. *Acta Psiquiátrica y Psicológica de América Latina*, 54(1), 55-61.
- 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.
- Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, 12, 269-275.
- 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.
- Montealegre, R. (2007). La solución de problemas cognitivos. Una reflexión cognitiva sociocultural. *Avances en Psicología Latinoamericana*, 25(2), 20-39.
- Organista Díaz, P. (2005). Conciencia y metacognición. *Avances en Psicología Latinoamericana*, 23, 77-89.
- Osborne, J. (1998). *Measuring metacognition: Validation of the Assessment of Cognition Monitoring Effectiveness*. Tesis doctoral inédita. New York: State University of New York at Buffalo.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology*, 29, 261-265.
- Schraw, G., Dunkle, M.E., Bendixen, L.D. & deBacker Roedel, T. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87(3), 433-444.
- Simmons, J.P. & Nelson, L.D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General*, 135(3), 409-428.

- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47, 143-148.
- Wallsten, T.S., Erev, I. & Budescu, D.V. (2000). The importance of theory: Response to Brenner (2000). *Psychological Review*, 107(4), 947-949.
- 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.

