

## **EFFECTO DEL ENTRENAMIENTO SOBRE LA CALIBRACIÓN SUBJETIVA DE ÉXITO EN TAREAS VERBALES<sup>1</sup>**

GUILLERMO MACBETH<sup>2</sup>

*Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina*

NURIA CORTADA DE KOHAN

*Universidad de Buenos Aires, Argentina*

### **ABSTRACT**

The calibration or confidence is broadly defined as the relation of correspondence or discrepancy between the subjective estimation of success and the actual success in a series of tasks. The underconfidence bias is defined as underestimation and the overconfidence bias as overestimation. Underconfidence and overconfidence biases are presented as cognitive phenomena. The main debiasing methods are analyzed. Two debiasing strategies are identified. The methodological strategy is differentiated from the cognitive strategy. An experiment that dissolves the underconfidence bias in verbal tasks through a simple training is presented and analyzed. It is suggested that the dissolution is due to training in calibration rather than training in the experimental task. The effect of adjustment in calibration is explained by the adaptive participation of metacognitive processing.

*Key words:* overconfidence bias, underconfidence bias, debiasing, calibration training, verbal performance.

---

1 Investigación financiada por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) de Argentina. Beca Postdoctoral concedida al primer autor. Agradecemos a Mg. Ing. María Itatí Gandulfo, Secretaria Académica de la Universidad Tecnológica Nacional, Facultad Regional Paraná, Argentina, por permitirnos aplicar las pruebas a sus alumnos de tecnología.

2 Correspondencia: GUILLERMO MACBETH, Dirección postal: IIPUS, Marcelo T. de Alvear 1312 (C1058AAV), Buenos Aires, Argentina. Correo electrónico: guillermo.macbeth@mail.salvador.edu.ar. NURIA CORTADA DE KOHAN. Correo electrónico: ncortada@psi.uba.ar.

## RESUMEN

Se define la calibración o confianza como la relación de correspondencia o discrepancia entre la estimación subjetiva de éxito y el éxito real obtenido en una serie de tareas. Se denomina sesgo de subconfianza a la subestimación y sesgo de sobreconfianza a la sobreestimación. Se analiza la estrategia metodológica y la estrategia cognitiva para la disolución de estos sesgos. Se presenta un experimento que disuelve el sesgo de subconfianza en tareas verbales mediante un entrenamiento simple. Se señala que la disolución se debe al entrenamiento en calibración antes que al entrenamiento en la tarea experimental. Se sugiere que el efecto de ajuste en la calibración se explica por la participación adaptativa de procesos metacognitivos.

*Palabras clave:* sesgo de sobreconfianza, sesgo de subconfianza, disolución de sesgos, entrenamiento en calibración, rendimiento verbal.

## INTRODUCCIÓN

Uno de los primeros estudios publicados sobre calibración subjetiva de éxito se realizó en el ámbito de la clínica psicológica. Hace más de cuatro décadas Oskamp (1965) propuso estudiar la confianza de un grupo de psicólogos clínicos respecto de sus propios diagnósticos. Con tal propósito aisló dos variables, una de rendimiento objetivo y otra de estimación subjetiva de éxito. Se operacionalizó la primera variable mediante el registro de la cantidad de pronósticos acertados, y la segunda por la seguridad o firmeza con que los psicólogos sostenían el acierto de sus pronósticos. Los resultados indicaron que la estimación subjetiva de éxito fue significativamente superior al éxito real obtenido. Este fenómeno se denominó *efecto o sesgo de sobreconfianza* y puede ser definido, en sentido amplio, como la sobreestimación del éxito subjetivo en comparación con el éxito real obtenido, para una serie de tareas específicas (Ko & Huang, 2007; Lichtenstein, Fischhoff & Phillips, 1982). El fenómeno opuesto a la sobreconfianza se denominó *efecto o sesgo de subconfianza*, esto es, la subestimación del éxito subjetivo en comparación con el éxito real (Kirchler & Maciejovsky, 2002). La *calibración o confianza* se define, en sentido amplio, como la relación de coincidencia o discrepancia que se observa entre la estimación subjetiva de éxito y el éxito real obtenido (Brenner, 2003; Fischhoff, 1982).

*Sesgos de la calibración*

Estos fenómenos se definen como distorsiones de la confianza o calibración subjetiva de éxito. Formalmente, se facilita la representación de estas variables y sus relaciones mediante la Ecuación 1:

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

La calibración o confianza subjetiva de éxito ( $C$ ) se obtiene de restar el éxito real u obtenido ( $O_i$ ) del éxito subjetivo o estimado ( $E_i$ ) para una serie de tareas específicas ( $i$ ). La sobreconfianza se observa cuando la suma que define a  $C$  genera puntajes positivos, es decir que el éxito estimado resulta mayor que el éxito obtenido ( $E > O$ ). La subconfianza, por el contrario, se observa cuando la  $C$  resulta negativa, es decir, el éxito estimado es menor que el obtenido ( $E < O$ ). El sujeto se encuentra bien calibrado cuando la diferencia entre  $E_i$  y  $O_i$  resulta igual a cero (la  $C$  no es positiva ni es negativa).

Estos sesgos de la calibración subjetiva de éxito se han observado para diversas tareas y diferentes poblaciones (Fernández-Duque & Black, 2007; Klayman, Soll, González-Vallejo & Barlas, 1999; Macbeth & Bogiaizian, 2007; Macbeth, Cortada de Kohan & Kohan Cortada, 2007; Macbeth, Razumiejczyk & Cortada de Kohan, 2006). Además de los estudios aportados por Oskamp (1965), que presentaron el

efecto de sobreconfianza en el diagnóstico psicológico, se cuenta con evidencia de distorsiones en campos tan diversos como la economía (Angner, 2006; Camerer & Lovallo, 1999; Dunne, Roberts & Samuelson, 1988; Malmendier & Tate, 2005; Shapiro & Khemani, 1987), la física (Henrion & Fischhoff, 1986), la política (Koheler, Brenner & Griffin, 2002), la educación (Clayson, 2005) y la medicina (Christensen-Szalanski & Bushyhead, 1981; Westbrook, Gosling & Coeira, 2005), entre otros (Svenson, 1981).

Estudios recientes sugieren que, bajo ciertas condiciones vinculadas a diferentes tareas y rendimientos de los participantes, se observan simultáneamente los efectos de sobreconfianza y subconfianza (Klayman *et al.*, 1999). La evidencia indica que los sujetos de bajo rendimiento tienden a sobreestimar su desempeño, mientras que los de alto rendimiento tienden a subestimarlos, en tareas de conocimientos generales (Macbeth, 2005). Se encontraron también diferencias entre sexos para este tipo de tareas. Los varones tienden a sobreestimar su rendimiento, mientras que las mujeres tienden a subestimarlos (Macbeth, Cortada de Kohan, Razumiejczyk & López Alonso, 2006).

#### *Disolución de los sesgos de la calibración*

Se ha intentado disolver estas distorsiones de la confianza o calibración subjetiva de éxito mediante dos estrategias, una metodológica o de artefactos estadísticos, otra cognitiva o de manipulación experimental de ciertos procesos de pensamiento. Cada estrategia ha desarrollado una colección de técnicas de eliminación de los sesgos.

La estrategia metodológica argumenta que las distorsiones de C son, posiblemente, artefactos generados por la manera en que se analizan los datos antes que fenómenos del pensamiento humano. En tal caso, la corrección metodológica de los errores de medición permitiría disolver los efectos espurios de sobreconfianza y subconfianza (Erev, Wallsten & Budescu, 1994; Gigerenzer,

Hoffrage & Kleinbölting, 1991). Los promotores de esta estrategia evitan afirmar que los sesgos de la calibración son sólo metodológicos (Brenner, 2000), pero le atribuyen buena parte de sus manifestaciones a diversos artefactos (Mellers, Hertwig & Kahneman, 2001). Sin embargo, la evidencia experimental generada bajo las restricciones sugeridas por estos autores muestra que la disolución de los efectos no siempre ocurre. Esto indica que los fenómenos bajo estudio son reales e incluso robustos (Brenner, 2003; Fischhoff, 1982) y que es conveniente estudiar su conformación mediante una estrategia cognitiva metodológicamente adecuada (Merkle & Van Zandt, 2006).

Una de las restricciones de mayor utilidad que promueven los defensores de la crítica metodológica se centra en el uso de *diseños representativos*, entendidos como el empleo de tareas experimentales homogéneas y confiables, compatibles con las condiciones de dificultad que presentan los contextos naturales (Gigerenzer *et al.*, 1991). Se exige un muestreo aleatorio no sólo de los participantes, sino también de las tareas experimentales. Esta objeción ecológica no ha encontrado respaldo sistemático en la evidencia empírica (García-Retamero & Dieckmann, 2006), sólo en algunos casos se ha eliminado el sesgo de sobreconfianza por el empleo de diseños representativos (Brenner, 2000; Koheler *et al.*, 2002).

Otra restricción metodológica de gran influencia es la que deriva del modelo estocástico de probabilidad subjetiva de Erev *et al.* (1994), que define formalmente la confianza ( $x$ ) como una función compuesta por la confianza verdadera ( $v_i$ ) y un componente de error ( $e_j$ ), para el evento  $i$  en el momento  $j$ , como se indica en la Ecuación 2.

$$x_{ij} = f(v_i, e_j) \quad (2)$$

El modelo incluye una definición primitiva de la regla de respuesta por la que esta confianza verdadera, contaminada por el error, genera la respuesta empírica  $y_{ij}$  de la Ecuación 3.

$$y_{ij} = g(x_{ij}) \quad (3)$$

La variación en la forma de las funciones de  $f$  y  $g$ , junto con la forma de la distribución del error  $e$ , permiten derivar diferentes casos a partir del modelo general de las Ecuaciones 2 y 3. Erev et al. (1994) demostraron que, en algunos casos derivados de este modelo general, se observa sobreconfianza o subconfianza en  $y_{ij}$  cuando la confianza verdadera ( $v_i$ ) se encuentra perfectamente calibrada ( $C = 0$  en Ecuación 1). Este modelo estocástico de probabilidad subjetiva presenta, sin embargo, severas limitaciones que fueron señaladas oportunamente por Brenner (2000, 2003; Wallsten, Erev & Budescu, 2000). Si bien la tradición psicométrica que se remonta a Thurstone (1927) ha hecho un irremplazable aporte al señalar la presencia de error estocástico en la medición de juicios psicológicos, tal como ocurre en el caso de la confianza (entendida como confianza verdadera más error), resulta objetable considerar al error verdadero  $v_i$  como más fundamental que la respuesta empírica  $y_{ij}$ . Sólo un diseño que permita discriminar empíricamente el error estocástico que contamina la respuesta empírica de la confianza verdadera permitiría tal prioridad. En cualquier otro caso resulta conveniente tratar la variable dependiente en toda su riqueza y complejidad (Brenner, 2000). La aplicación del modelo de Erev et al. (1994) permitiría desestimar los sesgos de la confianza como meros artefactos metodológicos en algunos diseños experimentales (Wallsten et al., 2000). Tal estrategia resulta, sin embargo, arbitraria porque aplica presupuestos que no pueden justificarse con los conocimientos psicológicos actualmente disponibles acerca de estos fenómenos.

Con un propósito diferente, Kahneman y Tversky (1982) han propuesto un índice de corrección general aplicable a las predicciones intuitivas (Simmons & Nelson, 2006). Estos autores ofrecen un procedimiento de corrección para el efecto de sobreconfianza en expertos. Este procedimiento propone reducir la discrepancia que genera la sobreestimación de  $E_i$  por sobre  $O_i$  en la Ecuación 1. Para ello definen el

concepto de predictibilidad  $p$ , entendido como la estimación subjetiva de la ocurrencia de un evento. La medida adecuada de  $p$  en un contexto de predicción lineal es la correlación entre las predicciones o estimaciones subjetivas ( $E_i$ ) y los resultados acertados o éxitos objetivos ( $O_i$ ). Es esta medida la que debe corregirse cuando se prevé la ocurrencia del sesgo de sobreconfianza. La corrección de  $p$  consiste en aproximar la estimación subjetiva ( $E_i$ ) a la estimación estadística que permite la consideración conjunta de otros casos similares o antecedentes ( $O_i$ ). La estimación corregida  $E_c$  se obtiene de sumar a la estimación estadística  $O_i$  la discrepancia entre  $E_i$  y  $O_i$ , es decir, entre la estimación subjetiva de éxito y el éxito real o estimación estadística ( $E_i - O_i$ ). El cómputo para la estimación corregida ( $E_c$ ) se presenta en la Ecuación 4.

$$E_c = \sum_{i=1}^n O_i + [\tau_i (E_i - O_i)] \quad (4)$$

A la estimación que puede obtenerse de los datos conocidos sobre casos similares ( $O_i$ ) se le suma un término que cuantifica la discrepancia entre la estimación subjetiva sobreconfiada y la estimación estadística ( $E_i - O_i$ ). Se multiplica este último término por el valor de  $\tau$ , que es un índice de calidad de la predictibilidad  $p$ . La Ecuación 5 muestra los cálculos necesarios para calcular la calidad de las estimaciones subjetivas de éxito, que en la Ecuación 4 se denomina  $\tau$ .

$$\tau = 2p - 1 \quad (5)$$

Cuando la  $p$  se cuantifica como probabilidad, la escala de predictibilidad para el caso de un evento que presenta sólo dos posibles resultados oscila entre 0,50 y 1 porque una  $p$  de 1/2 indica aleatoriedad. De esta manera, la variable  $\tau$  oscilaría entre 0 y 1 ( $0 \leq \tau \leq 1$ ). Luego, para cualquier valor de  $E_i$  superior a  $O_i$  se obtendrá una reducción de la sobreconfianza mediante el reemplazo de  $E_i$  por  $E_c$ , es decir, se reducirá la discrepancia entre ambas distribuciones. La Ecuación 6 presenta el formalismo de la confianza corregida  $C_c$  que se obtiene de reemplazar  $E$  por  $E_c$  en la Ecuación 1.

$$C_c = \sum_{i=1}^n (E_{C_i} - O_i) \quad (6)$$

La confianza corregida  $C_c$  de la Ecuación 6 será siempre menor que la confianza sin corregir  $C$  de la Ecuación 1 ( $C_c < C$ ). Esta estrategia metodológica presenta la ventaja de reducir la sobreconfianza en expertos, pero resulta insuficiente como método de disolución del sesgo, tal como lo explican Kahneman, Slovic y Tversky (1982).

La segunda estrategia de disolución de los sesgos de la confianza o calibración subjetiva de éxito no es metodológica sino cognitiva. Fischhoff (1982) ofrece un inventario de las técnicas clásicas que se han empleado para reducir los sesgos de sobreconfianza y subconfianza. Cada técnica presenta ventajas y limitaciones que se derivan de los presupuestos cognitivos que las guían. Fischhoff distingue entre manipulaciones de las tareas, de los participantes y de la relación entre ambos. Respecto de las tareas se presupone que la calibración es más exacta y, por tanto, menor es la ocurrencia de sesgos cuando, por ejemplo, mayor claridad presentan los ítems y menor sobrecarga cognitiva generan (Klayman *et al.*, 1999). Cuanto más difíciles son los ítems en tareas de conocimientos generales, mayor sobreconfianza se observa (Fischhoff, 1982; Gigerenzer *et al.*, 1991). En sentido inverso, cuanto más fáciles son las tareas, mayor subconfianza se observa (Macbeth *et al.*, 2006). Este fenómeno doble se denomina *efecto difícil-fácil* en las publicaciones especializadas (Juslin, Winman & Olsson, 2000; Lichtenstein *et al.*, 1982). Respecto de los participantes se presupone que calibran mejor cuando se les advierte sobre los sesgos habituales de la confianza, se les ofrece retroalimentación de los resultados o se los entrena (Gigerenzer *et al.*, 1991; Macbeth & Bogiaizian, 2007). Respecto de la relación entre tareas y participantes se presupone, por ejemplo, que los expertos calibran mejor que los no expertos, y que la educación (en sentido amplio) promueve una adecuada calibración (Angner, 2006; Armor & Taylor, 2002; Camerer & Lovallo, 1999).

La evidencia empírica acumulada sugiere que los sesgos de sobreconfianza y subconfianza son fenómenos robustos y, en general, resistentes a estas técnicas cognitivas de disolución (Fischhoff, 1982; Wilson, Centerbar & Brekke, 2002). Sólo el entrenamiento personalizado ha permitido reducir y, en algunos casos, eliminar el sesgo de sobreconfianza (Koheler *et al.*, 2002). El resto de las técnicas ha resultado, en general, ineficiente. Gigerenzer *et al.* (1991) han objetado, sin embargo, tanto la ubicuidad de la sobreconfianza (Nickerson, 1998) como la supuesta ineficiencia de las técnicas de disolución de los sesgos. Si bien los participantes expertos tienden a sobreestimar el éxito de su desempeño en tareas de su especialidad (Angner, 2006; Camerer & Lovallo, 1999; Westbrook *et al.*, 2005), se ha sugerido que el entrenamiento en tareas de calibración reduce e incluso disuelve el efecto de sobreconfianza (Fischhoff, 1982). Sin embargo, los participantes entrenados en diagnóstico clínico estudiados por Oskamp (1965) no redujeron ni eliminaron el sesgo de sobreconfianza al recibir nueva información relevante para la tarea diagnóstica. El análisis conjunto de estos datos sugiere que es necesario distinguir entre el entrenamiento en la tarea y el entrenamiento en calibración. Gigerenzer *et al.* (1991) señalan al respecto que se trata de dos procesos cognitivos complejos que presentan arquitecturas diferentes. El entrenamiento en la tarea es diferente del entrenamiento en la calibración de la tarea, aunque el segundo abarca al primero de manera recursiva. El entrenamiento en calibración subjetiva de éxito puede ser entendido como un proceso metacognitivo que monitorea y controla los procesos cognitivos activados por la tarea experimental (Fernández-Duque & Black, 2007). La eficacia de los procesos metacognitivos se diferencia de la eficacia de los procesos cognitivos, aunque tienden a la covariación sistemática (Macbeth, 2003, 2004).

Señalan Gigerenzer *et al.* (1991) que la advertencia explícita a los participantes acerca de los sesgos de sobreconfianza y subconfianza no logra reducir ni disolver los efectos. La recompensa económica tampoco logra mejorar la cali-

bración (Camerer & Lovallo, 1999). Estos hallazgos resultan coherentes con las revisiones ensayadas por Fischhoff (1982), Koheler *et al.* (2002) y Simmons & Nelson (2006).

#### *Calibración subjetiva de éxito en tareas verbales*

En tareas de tipo verbal, esto es, ítems de rendimiento experimental que no resultan diferentes del uso cotidiano del lenguaje, se espera buena calibración cuando las tareas experimentales son representativas respecto del universo de las tareas verbales que los participantes encuentran en su interacción habitual con el ambiente al que se encuentran adaptados (Gigerenzer *et al.*, 1991). Si las tareas experimentales, en cambio, no son representativas del ambiente cotidiano de los participantes, entonces se espera encontrar sobreconfianza o subconfianza, en función de la dificultad de los ítems de la prueba. Para una prueba difícil se espera observar sobreconfianza, para una fácil se espera observar subconfianza. Esto se explica, según Gigerenzer *et al.* (1991; García-Retamero & Dieckmann, 2006), por la activación de modelos mentales probabilísticos (MMP) almacenados en la memoria a largo plazo. Estos MMP son dispositivos cognitivos automáticos que se activan espontáneamente durante la toma de decisiones bajo incertidumbre cuando la estructura de la tarea experimental coincide con la estructura de algunas tareas respecto de las cuales el participante posee experiencia adaptativa. Se generan modelos mentales locales (MML) antes que MMP cuando la estimación subjetiva de éxito se aplica a eventos singulares antes que a series de eventos. Los MML son construcciones cognitivas que se ajustan a las restricciones de la tarea experimental específica, y consisten en procesos mnemónicos e inferenciales explícitos y simples, mientras que los MMP son patrones generales que se recuperan de la memoria a largo plazo por su correspondencia con los rasgos genéricos de la tarea experimental. De esta manera, los MMP generados para calibrar el éxito subjetivo en series de tareas verbales producen calibraciones subjetivas sesgadas en fun-

ción de la discrepancia entre la dificultad de la tarea experimental y la dificultad de tareas verbales habituales y conocidas, de las que se almacena en la memoria a largo plazo un registro de la frecuencia de éxito. Si las tareas experimentales resultan difíciles, se espera sobreconfianza, y si son fáciles se espera subconfianza, para los casos en que la prueba experimental no sea representativa del ambiente natural de los participantes. La subconfianza se entiende, en este contexto, como el resultado de una calibración que se estima en función de un universo de tareas cuya media de dificultad es mayor que la media de dificultad de la muestra experimental seleccionada. Es decir, las tareas experimentales son más difíciles que las tareas habituales, y por ello se genera un rendimiento objetivo más bien bajo. La sobreconfianza, en sentido inverso, se entiende como la discrepancia entre tareas experimentales que resultan demasiado fáciles en comparación con las tareas del universo conocido, por lo cual se generan niveles altos de rendimiento. La calibración subjetiva de éxito para series de tareas no se genera sobre la base del rendimiento real, sino a partir de un MMP estable. Luego, los sujetos de bajo rendimiento tienden a incurrir en el sesgo de sobreconfianza y los de alto rendimiento son propensos al sesgo de subconfianza (Macbeth, 2005; Macbeth *et al.*, 2006).

Se presenta a continuación un experimento que pone a prueba varias hipótesis relacionadas con el efecto del entrenamiento sobre la calibración subjetiva de éxito en tareas verbales. Para operacionalizar la estimación subjetiva de éxito  $E$  se ha consignado a los participantes la tarea de responder cuántas de las respuestas dadas en una prueba de rendimiento verbal son correctas, es decir, se ha evaluado la calibración  $C$  para una serie de tareas antes que para eventos singulares. Este diseño se ubica en el contexto de lo que Gigerenzer *et al.* (1991) denominan MMP.

La Hipótesis 1 ( $H_1$ ) afirma que los participantes sin entrenamiento presentan el sesgo de subconfianza ( $E < O$ ) en tareas verbales fáciles. Si las tareas experimentales resultan difíciles, se

pronostica el sesgo de sobreconfianza ( $E > O$ ). La Hipótesis 2 ( $H_2$ ) sostiene que el entrenamiento disuelve el sesgo de subconfianza (o de sobreconfianza, para el caso de tareas fáciles) afirmado en la  $H_1$ . La Hipótesis 3 ( $H_3$ ) sostiene que la disolución del sesgo de subconfianza (o de sobreconfianza) se debe al entrenamiento en calibración antes que al entrenamiento en la tarea experimental.

La  $H_1$  surge de la teoría de los MMP, propios de la estimación subjetiva de éxito para series de eventos. Los MML, en cambio, son propios de la estimación subjetiva de éxito para eventos singulares. La predicción de la  $H_1$  se ajusta a la teoría ecológica o adaptativa de la confianza propuesta por Gigerenzer *et al.* (1991; García-Retamero & Dieckmann, 2006).

La  $H_2$  surge de la conjetura que afirma la eficacia metacognitiva de la calibración, entendida como el ajuste adaptativo permanente de las expectativas subjetivas que proveen el monitoreo y el control recursivos que acompañan a los procesos cognitivos en general (Hoffrage, Hertwig & Gigerenzer, 2000; Macbeth, 2003, 2004). La  $H_2$  resulta, además, coherente con estudios anteriores (Fischhoff, 1982; Gigerenzer *et al.*, 1991) que señalan la eficacia del entrenamiento personalizado. Se espera que el entrenamiento en calibración subjetiva de éxito disuelva el sesgo de subconfianza (o de sobreconfianza, según el caso) por la actualización o ajuste del MMP a las características específicas de las tareas experimentales.

La  $H_3$  apoya la distinción propuesta por Gigerenzer *et al.* (1991) que diferencia los procesos cognitivos de la tarea experimental de los procesos cognitivos y metacognitivos de la calibración de éxito en tareas experimentales. Para poner a prueba esta hipótesis se recomienda una comparación intra-grupo antes que inter-grupos (Mellers *et al.*, 2001). Al comparar cada participante consigo mismo se logra mantener controladas ciertas variables que resultan difíciles de controlar en comparaciones inter-grupos como son, por ejemplo, la experticia en la tarea espe-

cífica o las diferencias de calidad del monitoreo metacognitivo de los participantes (Kahneman, 2003).

## MÉTODO

### *Participantes*

Participaron en este experimento 159 estudiantes universitarios argentinos de grado y postgrado de dos universidades, una nacional (Universidad Tecnológica Nacional) y otra privada (Universidad del Salvador). Predominaron los varones ( $n_v = 102$ ; 64,2%) sobre las mujeres ( $n_m = 57$ ; 35,8%). La edad promedio de los participantes resultó de 27,78 años ( $de = 8,098$ ). Se establecieron dos grupos, uno experimental ( $n_e = 78$ ) y otro control ( $n_c = 81$ ). Se mantuvieron proporciones similares de varones y mujeres en ambos grupos. La asignación de los participantes a cada grupo fue aleatoria.

### *Instrumento*

Se administró en ambos grupos el test BAIRES de rendimiento verbal (Cortada de Kohan, 2003) que mide el desempeño de estudiantes universitarios de habla hispana en tareas de sinónimos y definiciones. El test BAIRES es un instrumento de papel y lápiz que presenta adecuadas propiedades psicométricas de confiabilidad y validez, que se mantuvieron dentro de lo esperado para los participantes del presente estudio ( $\alpha$  de Cronbach = 0,699). 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. Para la construcción del test BAIRES se aplicaron los métodos de la Teoría de la Respuesta al Ítem (Cortada de Kohan, 1999). La versión extensa del instrumento que se aplicó en ambos grupos 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 test (Cortada de Kohan, 2003) se aplicó, a modo de entrenamiento, a los participantes del grupo experimental. La versión abreviada del test BAIREs está compuesta por 16 ítems, de los cuales 8 son tareas de definiciones y 8 son tareas de sinónimos. 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.

La aplicación del test BAIREs permitió medir el éxito real obtenido ( $O$ ) por los participantes en tareas verbales mediante un instrumento válido y confiable. Junto con esta variable de rendimiento objetivo se midió la estimación subjetiva de éxito ( $E$ ) para la serie de tareas verbales respondidas mediante una única estimación global de aciertos. Se solicitó a los participantes que estimen cuántas de sus respuestas son correctas sobre un total de 16 para la versión abreviada, aplicada sólo al grupo experimental, y sobre un total de 34 para la versión extensa, aplicada en ambos grupos.

#### *Procedimiento*

Al grupo control ( $n_c = 81$ ) se le administró sólo la versión extensa del test BAIREs. Al grupo experimental ( $n_e = 78$ ) se le administró la versión abreviada del test BAIREs en la primera fase del estudio, a modo de entrenamiento, y la versión extensa del mismo test en la segunda fase del estudio, para evaluar el efecto del entrenamiento sobre la calibración. Luego de completar la primera fase, que incluyó la tarea de responder todas las preguntas ( $O$ ) junto con la tarea de estimación subjetiva de éxito ( $E$ ), los participantes del grupo experimental recibieron la consigna de evaluar su propia calibración ( $C$ ). El experimentador indicó las respuestas correctas y solicitó que consideren la exactitud de su

calibración subjetiva de éxito, aunque sin aclaraciones adicionales. Una vez concluido este entrenamiento, se solicitó la devolución de los protocolos y se entregó a cada participante la versión extensa del test BAIREs, en el que se consignaban tareas idénticas a las de la fase de entrenamiento. No se implementaron tareas distractoras entre la fase de entrenamiento y la fase de evaluación.

La administración del instrumento se realizó en forma grupal. Se realizaron en total 10 sesiones para la recolección de datos ( $10 = n = 25$ ), en condiciones experimentales homogéneas. Los participantes de ambos grupos completaron la versión extensa del BAIREs en aproximadamente 20 minutos. La versión abreviada aplicada al grupo experimental en fase de entrenamiento no consumió más de 10 minutos en ninguna sesión.

## RESULTADOS

### *El sesgo de subconfianza en tareas de conocimientos generales ( $H_1$ )*

Los participantes del grupo control ( $n_c = 81$ ) incurrieron en el sesgo de subconfianza ( $E < O$ ). La media de aciertos o éxitos reales observados ( $O$ ) resultó de 19,74 ( $de = 4,089$ ) mientras que la media de éxitos estimados ( $E$ ) resultó de 17,53 ( $de = 5,795$ ), sobre un total de 34 ítems. La distribución de ambas variables resultó compatible con la presunción de normalidad e igualdad de varianzas por las pruebas de Kolmogorov-Smirnov y Levene, respectivamente. La prueba  $t$  para muestras apareadas indicó un predominio significativo de  $O$  por sobre  $E$  ( $t = -3,816$ ;  $p < 0,01$ ) para los participantes del grupo control. También los participantes del grupo experimental ( $n_e = 78$ ) subconfiaron en la fase de entrenamiento. La variable  $O$  registró en este grupo una media de 14,15 ( $de = 2,039$ ) y una mediana de 15 ( $dam = 0,231$ ), mientras que la variable  $E$  obtuvo una media de 12,40 ( $de = 2,741$ ) y una mediana de 13 ( $dam = 0,310$ ), sobre un total de 16 ítems. Este puntaje de  $O$  para el grupo expe-



rimental en fase de entrenamiento representa una probabilidad de acierto  $P_{(a)}$  de 0,906 (en función de la media, y de 0,937 en función de la mediana), es decir que se trata de un rendimiento alto ante una prueba relativamente fácil, por lo cual se espera encontrar subconfianza. La facilidad de las tareas generó una distribución asimétrica de  $O$ . Tanto la  $E$  como la  $O$  presentan asimetría a izquierda, es decir, una tendencia central a los puntajes altos. Se compararon las diferencias entre ambas variables mediante la prueba de Wilcoxon. Se encontró una diferencia significativa ( $p < 0,01$ ) que indica la ocurrencia del sesgo de subconfianza por el predominio de  $O$  sobre  $E$ . De los 78 participantes, 54 obtuvieron subconfianza ( $E < O$ ), 5 obtuvieron sobreconfianza ( $E > O$ ) y 19 obtuvieron una perfecta calibración ( $E = O$ ).

En síntesis, los participantes no entrenados en tareas de calibración, tanto del grupo experimental (en fase de entrenamiento) como del grupo control, incurrieron en el sesgo de subconfianza. Se cumple en este estudio el pronóstico vinculado a la facilidad relativa de la tarea. Si bien el rendimiento en el grupo control, con la versión extensa ( $P_{(a)} = 0,58$ ), no resultó tan alto como en el grupo experimental, con la versión abreviada ( $P_{(a)} = 0,90$ ), se puede afirmar que en ambos casos las tareas permiten lograr niveles altos o moderadamente altos de éxitos o aciertos. La  $H_1$  resultó coherente con la evidencia, es decir, los participantes sin entrenamiento presentan el sesgo de subconfianza ( $E < O$ ) en tareas verbales fáciles.

#### *El entrenamiento disuelve el sesgo de subconfianza ( $H_2$ )*

La  $H_2$  sostiene que el entrenamiento disuelve el sesgo de subconfianza (o sobreconfianza) señalado en la  $H_1$ . Se estudiaron las variables  $E$  y  $O$  mediante las respuestas a la versión extensa del BAIREs dadas por los participantes del grupo experimental, luego de haber sido entrenados con la versión abreviada del mismo instrumento. Ambas variables resultaron normales y homocedásticas por las pruebas de Kolmogorov-Smirnov y Levene,

respectivamente. El puntaje de  $E$  ( $media_E = 22,04$ ;  $de_E = 6,212$ ) resultó levemente superior al de  $O$  ( $media_O = 21,68$ ;  $de_O = 5,221$ ). La diferencia entre  $E$  y  $O$  resultó no significativa por la prueba  $t$  para muestras apareadas ( $t = 0,507$ ;  $p = 0,614$ ).

En síntesis, la  $H_2$  resultó compatible con la evidencia experimental. El sesgo de subconfianza observado en participantes no entrenados en calibración subjetiva de éxito para tareas verbales se disolvió con un entrenamiento simple, implementado mediante un ejercicio de calibración sobre una versión abreviada de la misma prueba. Si bien la comparación inter-grupos resulta desaconsejable para este tipo de fenómenos debido a la variabilidad que generan algunas variables críticas aún no identificadas (Mellers *et al.*, 2001), nótese que la sobreconfianza se encontró en participantes no entrenados, tanto del grupo control como del grupo experimental en fase de entrenamiento, mientras que el efecto desapareció en el grupo experimental luego del entrenamiento.

#### *La disolución del sesgo de subconfianza se debe al entrenamiento en calibración ( $H_3$ )*

La forma adecuada de poner a prueba esta hipótesis consiste en comparar el comportamiento de la variable  $O$  en el grupo experimental, antes y después del entrenamiento. Si los éxitos reales obtenidos por los participantes no se incrementaron significativamente, entonces el entrenamiento modificó los procesos cognitivos de la calibración o confianza antes que los de la tarea experimental específica, esto es, del rendimiento verbal. Se transformaron los puntajes brutos de  $O$  registrados antes y después del entrenamiento en puntajes  $z$  debido a que la versión abreviada del BAIREs presenta una escala de 16 puntos, mientras que la versión extensa presenta 34 puntos. Estas transformaciones a puntajes  $z$  permiten comparar el comportamiento de  $O$  antes y después del entrenamiento sobre la base de una métrica común. La distribución de  $z_O$  antes del entrenamiento resulta, al igual que la distribución de  $O$ , asimétrica a izquierda e incompatible con la presunción de normalidad por la prueba de Kolmogorov-Smirnov ( $Z = 1,886$ ;  $p = 0,002$ ).

La diferencia entre ambas  $z_o$  resultó no significativa por la prueba de Wilcoxon ( $p = 0,083$ ). De los 78 participantes, 45 mejoraron el rendimiento luego del entrenamiento ( $z_{Oantes} < z_{Odespués}$ ), 33 lo empeoraron ( $z_{Oantes} > z_{Odespués}$ ), y ninguno mantuvo un rendimiento proporcional estable ( $z_{Oantes} = z_{Odespués}$ ).

En síntesis, la  $H_3$  sostiene que la disolución del sesgo de subconfianza se debe al entrenamiento en calibración antes que al entrenamiento en la tarea experimental. La evidencia indica que el entrenamiento disolvió el sesgo de subconfianza pero no logró mejorar el rendimiento en tareas verbales homogéneas. La  $H_3$  resultó compatible con esta evidencia.

## DISCUSIÓN

Los sesgos de sobreconfianza y subconfianza pueden ser entendidos, en términos generales, como distorsiones en la calibración subjetiva de éxito en la toma de decisiones bajo incertidumbre. La sobreconfianza consiste en sobreestimar el éxito subjetivo en comparación con el éxito objetivo para una serie de tareas, mientras que la subconfianza es el fenómeno inverso, la subestimación. La disolución de estos sesgos o efectos es uno de los problemas clásicos de investigación que figuran en la agenda de los dos programas dominantes de la psicología del pensamiento bajo incertidumbre. El programa de Sesgos y Heurísticos (Cortada de Kohan & Macbeth, 2004; Gilovich, Griffin & Kahneman, 2002; Kahneman *et al.*, 1982; Kahneman & Tversky, 2000) y el programa de Racionalidad Ecológica (García-Retamero & Dieckmann, 2006; García-Retamero, Ramos & Catena, 2008; Gigerenzer, 2000; Gigerenzer & Selten, 2002; Gigerenzer, Todd & ABC Research Group, 1999) han logrado algunos avances que no se encuentran unificados aún. Resulta notable la falta de estudios meta-analíticos sobre sesgos y heurísticos en la bibliografía especializada, más allá de la acumulación de experimentos individuales similares que estudian, por ejemplo, los sesgos de sobreconfianza y subconfianza. Las revisiones

disponibles sobre el estado de la cuestión sugieren que se trata de fenómenos robustos y sistemáticos, cuya variabilidad puede explicarse, posiblemente, por la interacción de variables de la tarea (por ejemplo, su dificultad) con variables de los participantes (por ejemplo, rendimiento, sexo) y de la interacción entre la tarea y los participantes (por ejemplo, experticia). Se requiere, además, la implementación de métodos que eviten la generación de efectos espurios como los señalados por Erev *et al.* (1994) y luego discutidos por Brenner (2000). Se destaca la estrategia metodológica de reducción del sesgo de sobreconfianza propuesta por Kahneman & Tversky (1982), aunque su utilidad se reduce a la moderación del efecto y no a su disolución. Por su parte, la teoría ecológica de los modelos mentales locales (MML) y probabilísticos (MMP) sugiere que el entrenamiento para la disolución de los sesgos de la confianza debe centrarse en la calibración antes que en la tarea experimental específica. La evidencia hallada en el presente experimento resulta coherente con estos antecedentes. La subconfianza observada en participantes sin entrenamiento en tareas fáciles se disolvió luego del entrenamiento. Se encontró que el entrenamiento no mejoró el rendimiento en la tarea, pero logró generar una discrepancia no significativa entre la estimación subjetiva y el éxito objetivo en una prueba de rendimiento verbal. No se ha encontrado en la bibliografía especializada ningún antecedente de tal comparación, por lo que se considera como un aporte original del presente estudio. Se propone aquí que el ajuste en la calibración se debe a la participación de procesos metacognitivos de monitoreo y control de las tareas experimentales. Es posible que la activación de un modelo mental probabilístico correspondiente con la tarea experimental propuesta se haya actualizado espontáneamente con la información provista por el entrenador en la primera fase del experimento sobre la calidad de la calibración.

Se concluye que el entrenamiento en calibración subjetiva de éxito disuelve el sesgo de subconfianza en tareas verbales fáciles. Se pronostica la misma disolución para los casos

de sobreconfianza generados por tareas verbales difíciles. Se sugiere que el proceso de corrección de la calibración observado en este estudio para tareas verbales se produce tam-

bién en otras tareas habituales para los participantes, siempre que el entrenamiento se centre en la calibración antes que en la tarea experimental.

#### 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.
- 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.
- 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.
- Cortada de Kohan, N. (1999). *Teorías Psicométricas y Construcción de Tests*. Buenos Aires: Lugar Editorial.
- Cortada de Kohan, N. (2003). *BAIRES. Test de Aptitud Verbal*. Madrid: TEA.
- Cortada de Kohan, N. & Macbeth, G. (2004). El reconocimiento público a la Psicología Cognitiva (A Propósito del Premio Nobel Otorgado en 2002 a Daniel Kahneman). *Irice*, 18, 159-164.
- Dunne, T., Roberts, M.J. & Samuelson, L. (1988). Patterns of firm entry and exit in U.S. manufacturing industries. *RAND Journal of Economics*, 19, 495-515.
- 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. & Dieckmann, A. (2006). Una visión crítica de los heurísticos rápidos y frugales. *Revista Latinoamericana de Psicología*, 38(3), 509-522.
- García-Retamero, R., Ramos, M. & Catena, A. (2008). El procesamiento de compuestos estímulares en juicios de causalidad. *Revista Latinoamericana de Psicología*, 40(1), pp.
- Gigerenzer, G. (2000). *Adaptive Thinking. Rationality in the Real World*. Oxford: Oxford University Press.
- Gigerenzer, G. & Selten, R. (Eds.) (2002). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, MA: The MIT Press.
- Gigerenzer, G., Todd, P.M. & ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.
- 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.
- Henrion, M. & Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics*, 54, 791-797.
- Hoffrage, U., Hertwig, R. & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(3), 566-581.
- 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.
- Kirchler, E. & Maciejovsky, B. (2002). Simultaneous over- and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty*, 25(1), 65-85.
- 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.
- Ko, K.J. & Huang, Z. (2007). Arrogance can be a virtue: Overconfidence, information acquisition, and market efficiency. *Journal of Financial Economics*, 84(2), 529-560.
- Koheler, 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. (2003). Metacognición y atribución de valor. Eficacia del monitoreo metacognitivo en procesos de atribución de verdad, falsedad y novedad. Tesis doctoral inédita. Buenos Aires: Universidad del Salvador.
- Macbeth, G. (2004). Metacognition and value attribution. *The Brunswik Society Newsletter*, November, 19.
- Macbeth, G. (2005). Confidence calibration and performance differences in general knowledge tasks. *The Brunswik Society Newsletter*, 20, 11.
- 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. & Kohan Cortada, A. (2007). Efecto de la experticia matemática en el sesgo de sobreconfianza. *Perspectivas en Psicología*, 4(1), 13-18.
- Macbeth, G. Razumiejczyk, E. & Cortada de Kohan, N. (2006). El sesgo de sobreconfianza en tareas verbales y matemáticas. *Investigaciones en Psicología. Revista del Instituto de Investigaciones de la Facultad de Psicología, UBA*, 11(3), 47-58.
- 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.
- Malmendier, U. & Tate, G.A. (2005). Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management*, 11(5), 649-659.
- 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.
- Nickerson, R.S. (1998). Confirmation bias: An ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology*, 29, 261-265.
- Shapiro, D. & Khemani, R.S. (1987). The determinants of entry and exit reconsidered. *International Journal of Industrial Organization*, 5, 15-26.
- 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.
- Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Wallsten, T.S., Erev, I. & Budescu, D.V. (2000). The importance of theory: Response to Brenner (2000). *Psychological Review*, 107(4), 947-949.
- Westbrook, J.I., Gosling, A.S. & Coeira, E.W. (2005). The impact of an online evidence system on confidence in decision making in a controlled setting. *Medical Decision Making*, 25(2), 178-185.
- 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.

Recepción: septiembre de 2006

Aceptación final: noviembre de 2007.