

# PAPELES DEL PSICÓLOGO

EVALUAR LA SIMULACIÓN, EL FINGIMIENTO Y LA MENTIRA



---

DETECCIÓN DE LA SIMULACIÓN EN LA EVALUACIÓN PSICOLÓGICA  
FINGIMIENTO EN LOS ÁMBITOS FORENSE, CLÍNICO Y LABORAL

---

## Sección monográfica

- 57. INTRODUCCIÓN:** Simulación, engaño y mentira.  
*Serafín Lemos Giráldez*
- 59.** Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG).  
*Ramón Arce y Francisca Fariña*
- 78.** ¿Se pillan antes a un mentiroso que a un cojo? Sabiduría popular frente a conocimiento científico sobre la detección no-verbal del engaño.  
*Jaume Masip*
- 92.** El análisis de contenido basado en criterios (CBCA) en la evaluación de la credibilidad del testimonio.  
*Verónica Godoy-Cervera, Lorenzo Higuera*
- 99.** La simulación de enfermedad física o trastorno mental.  
*Mercedes Inda Caro, Serafín Lemos Giráldez, Ana María López Rodrigo y José Luis Alonso Rionda*
- 109.** El engaño y la mentira en los trastornos psicológicos y sus tratamientos.  
*Manuel Porcel Medina y Rubén González Fernández*
- 115.** Personalidad y deseabilidad social en contextos organizacionales: Implicaciones para la práctica de la Psicología del trabajo y las organizaciones.  
*Jesús F. Salgado*
- 129.** Evaluación de las distorsiones de respuestas mediante el MMPI-2.  
*Héctor González Ordi e Iciar Iruarizaga Díez*

## Special Section

- 57. INTRODUCTION:** Malingering, deception and lie.  
*Serafín Lemos Giráldez*
- 59.** Psychological evidence in court on statement credibility, psychological injury and malingering: The Global Evaluation System (GES).  
*Ramón Arce y Francisca Fariña*
- 78.** Is the liar sooner caught than the cripple? Popular wisdom versus scientific knowledge about nonverbal detection of deception.  
*Jaume Masip*
- 92.** Criterial based content analysis (CBCA) on the statement credibility assesment.  
*Verónica Godoy-Cervera, Lorenzo Higuera*
- 99.** Malingering of organic illness or mental disorder.  
*Mercedes Inda Caro, Serafín Lemos Giráldez, Ana María López Rodrigo y José Luis Alonso Rionda*
- 109.** Deception and lie in psychological disorders and its treatments.  
*Manuel Porcel Medina y Rubén González Fernández*
- 115.** Personality and social desirability in organizational settings: Implications for the practice of work and organizational psychology.  
*Jesús F. Salgado*
- 129.** Assessment of self-report distortions in the MMPI-2.  
*Héctor González Ordi e Iciar Iruarizaga Díez*

## PAPELES DEL PSICOLOGO

### Edita

Colegio Oficial de Psicólogos, España

### Consejo Editorial

Junta de Gobierno del Colegio Oficial de Psicólogos

### Decano

Francisco J. Santolaya Ochando

### Vicedecano 1

Eduardo Montes Velasco

### Vicedecano 2

Manuel Berdullas Temes

### Secretario

Manuel Mariano Vera Martínez

### Vicesecretario

Julián Baltasar Jaume

### Tesorero

Juan Manuel Egurtza Muniain

### Vocales

Rosario Carcas, Fernando Chacón, Aurora Gil Álvarez, José Ramón Fernández Hermida, Lorenzo Gil, Rosa Jiménez Tornero, Margarita Laviana, Cristina López Díaz, Manuel Enrique Medina, Manuel Rodríguez Fernández y Concepción Santo Tomás de Abajo.

### Consejo Asesor

José Antonio Aldaz, Esteban Alonso, Isaac Amigo, José Arévalo Serrano, Pilar Arránz, José María Arredondo, Dositeo Artiaga, M<sup>a</sup> Dolores Avia, Sabino Ayestarán, Francisco Bas, Vicent Bermejo, Amalio Blanco, Cristina Botella, Carmen Bragado, Gualberto Buela, José Buendía, Vicente Caballo, Francisco Cabello, José Cáceres, Rosa Calvo, Fernando Calvo, Amalia Cañas, Antonio Capafons, José Carlos Caracuel, Helio Carpintero, Mario Carretero, José Antonio Carobles, Miguel Costa, Antonio Coy, Sixto Cubo, Piedad Cueto, Fernando Díaz Albo, María José Díaz-Aguado, Jesús A. De Diego, Raúl De Diego, Andrés Duarte López, Rocío Fernández Ballesteros, Nicolás Fernández Losa, Jorge Fernández Del Valle, Concepción Fernández Rodríguez, Alfredo Fornos, Enrique García Huete, Miguel Anxo García Álvarez, César Gilolmo, Jesús Gómez Amor, Jorge L. González Fernández, Julio Antonio González García, José Gutiérrez Terrazas, Adolfo Hernández Gordillo, Florencio Jiménez Burillo, Cristóbal Jiménez Jiménez, Annette T.Kreuz, Francisco Javier Labrador, José Carlos León Jarriego, Jesús Ramón Loitegui, Roberto Longhi, Aquilino Lousa, Araceli Maciá, Emiliano Martín, María Angeles Martínez Esteban, José Joaquín Mira, Luis Montoro, José Muñiz, Nicomedes Naranjo, Conrado Navalón, José Ignacio Navarro Guzmán, Luis De Nicolás, Soledad Ortega Cuenca, Pedro Pérez García, Marino Pérez Álvarez, Félix Pérez Quintana, José Luis Pinillos, José Antonio Portellano, José María Prieto, Ismael Quintanilla, Francisco Ramos, Jesús Rodríguez Marín, Carlos Rodríguez Sutil, José Ignacio Rubio, Carlos Samaniego, Aurelia Sánchez Navarro, Javier Urra, Miguel Angel Vallejo y Jaime Vila.

### Director

Serafín Lemos Giráldez

### Directores asociados

José Ramón Fernández Hermida, Manuel Enrique Medina Tornero, José Carlos Núñez Pérez y José María Peiró Silla

### Diseño y Maquetación

Juan Antonio Pez Martínez

### Redacción, administración y publicidad

Juan Antonio Pez Martínez  
Consejo General de Colegios Oficiales de Psicólogos  
C/ Conde de Peñalver, 45-5<sup>o</sup> Izq.  
28006 Madrid - España  
Tels.: 91 444 90 20 - Fax: 91 309 56 15  
E-mail: papeles@correo.cop.es

### Impresión

Intigraf S.L.  
C/ Cormoranes, 14. Polígono Industrial La Estación.  
28320 Pinto Madrid

### Depósito Legal

M-27453-1981 / ISSN 0214-7823

De este número 92 de Papeles del Psicólogo se han editado 46.600 ejemplares.  
Los editores no se hacen responsables de las opiniones vertidas en los artículos publicados.

Papeles del Psicólogo está incluida en Psycodoc y en las bases de datos del ISOC (Psedisoc) y en IBECS. También se puede consultar en la página WEB del Colegio de Psicólogos:

<http://www.cop.es>



## SIMULACIÓN, ENGAÑO Y MENTIRA MALINGERING, DECEPTION AND LIES

**Serafín Lemos Giráldez**

*Facultad de Psicología. Universidad de Oviedo*

**E**s evidente que la simulación, el engaño y la mentira forman parte de la naturaleza humana y están presentes tanto en la esfera personal como en la vida social. Desde los más breves saludos estructurados, del estilo de "Buenos días, ¿qué tal estás? Bien, gracias", en donde las palabras pierden su significado genuino para dar paso a meras fórmulas de cortesía, hasta las formas más elaboradas de comunicación en el complejo entramado social se apoyan en un juego de roles, en donde se entremezcla lo que la persona es con lo que aparenta ser, la realidad con la imagen, la función de autor con la de actor.

El engaño no es exclusivo de la especie humana sino que es también una característica que está presente en los primates y en otros animales que viven en entornos sociales de gran complejidad; y en los reinos animal y vegetal son numerosos los seres vivos que han desarrollado, en el proceso evolutivo, capacidades de camuflaje y de adaptación muy elaboradas, que han prosperado gracias al efecto de confundir a sus competidores o a sus depredadores.

La supervivencia en un medio social complejo ha favorecido el desarrollo de la neocorteza cerebral en el ser humano y en otros mamíferos superiores, que ha hecho posible la adquisición de habilidades mentales extraordinarias, como son la auto-conciencia y la teoría de la mente, que permiten no sólo reconocer características personales sino también anticipar los pensamientos e intenciones de los congéneres, aumentando así las habilidades sociales y la cohesión grupal. Sin entrar a discutir si el primate es consciente de su conducta, si su pensamiento es reflexivo o si sus actos obedecen a un plan preconcebido (hipótesis mentalista), o bien si dichas habilidades son resultado de meros reflejos innatos o de procesos de aprendizaje instrumental (hipótesis conductista), lo que parece evidente es que estas capacidades están en el origen de lo que Whiten y Byrne (1997) han denominado "inteligencia maquiavélica"; teoría según la cual los primates han conseguido desarrollar diversas estrategias sociales que son ventajosas para la supervivencia, recurriendo al uso convenenciero de comportamientos agonistas o cooperativos, según lo exija la situación. La inteligencia maquiavélica es una capacidad que parece haber sido inducida por la necesidad de dominar formas cada vez más refinadas de manipulación y defraudación en el medio social, y que se manifiesta a través del uso de estrategias de disimulo, mentira y engaño tácticos. Como también señala Smith (2005), la inteligencia maquiavélica

pudo haber sido el motor que empujó a nuestros antepasados a ir adquiriendo cada vez mayor inteligencia y a hacerse cada vez más aficionados a mudar de opinión, a cerrar tratos, a farolear y a confabularse con otros; por lo que estima que los seres humanos son mentirosos natos, habiendo desarrollado formas mucho más sofisticadas de disimulo que nuestros parientes primates más cercanos.

Pero el disimulo y el engaño a los demás no hubiera alcanzado tales órdenes de magnitud si los humanos no hubiéramos desarrollado también la habilidad de engañarnos a nosotros mismos. El autoengaño nos ayuda a mentir a otros de manera más convincente, y la capacidad para creernos nuestras propias mentiras nos ayuda a embaucar más eficazmente a los demás. Por otra parte, nos permite alcanzar el grado de perfección de "mentir con sinceridad", sin necesidad de hacer un montaje teatral para fingir que estamos diciendo la verdad. Esta es la tesis del sociobiólogo Robert Trivers (2002), quien sostiene que la función capital del autoengaño es poder engañar más fácilmente a otros, por cuanto la credulidad en el propio cuento lo hace más convincente para los demás.

Así pues, el disimulo, la mentira implícita o el engaño deliberado forman parte de todos los escenarios en los que transcurre la vida social humana. En un proceso evolutivo cuyas etapas se van consumiendo desde la infancia, se va perdiendo la espontaneidad conforme se asienta la convicción de que la sinceridad no siempre es posible ni conveniente, porque puede perjudicar a la otra persona o a uno mismo. Por eso, mienten los amigos bien intencionados con el fin de halagar, de edulcorar la realidad, de dar apoyo o de proteger a la persona estimada; mienten los gobernantes y los líderes sociales para conseguir sus propósitos, para evitar problemas o para seducir al electorado (llegando a la paradoja de que son los más mentirosos quienes muestran el mayor empeño en desenmascarar las mentiras de sus adversarios); mienten los medios de comunicación, ocultando información o publicando información interesada, enfatizando noticias o contrarrestándolas con otras; mienten los publicistas y los vendedores en todas las transacciones comerciales para persuadir a sus clientes; y, entre otros, mienten también los profesionales para defender sus intereses, el reconocimiento social o para lograr la satisfacción de sus clientes. En definitiva, todas las personas intentan acomodar la realidad a sus propias intenciones, expectativas o necesidades; pero lo sorprendente es que, a sabien-



das de que el mundo es así, actuamos como si todo fuera verdad o tal vez necesitamos persuadirnos de que lo es.

En los diferentes ámbitos del ejercicio profesional del psicólogo, el disimulo, el encubrimiento, la exageración, el filtrado y el falseo de la información suministrada o la negación de problemas son fenómenos muy frecuentes, y constituyen un importante obstáculo a la hora de realizar una correcta evaluación de cada caso y de tomar decisiones. Con toda seguridad, los mismos inconvenientes existen en cualquier otra actividad profesional; pero en nuestro medio, sin embargo, dichos comportamientos suelen estar determinados por diversas causas, a veces patológicas (la existencia de algún trastorno mental), criminológicas (la necesidad de evitar responsabilidades jurídicas) o meramente adaptativas (conseguir determinados objetivos en circunstancias adversas) (Rogers, 1997).

Una premisa imprescindible para el ejercicio profesional del psicólogo es la cooperación y la honestidad del cliente o paciente; de tal manera que la correcta evaluación y el diagnóstico psicológicos dependen de la honestidad y del deseo de aportar información, por la persona evaluada, así como de la precisión y veracidad de los datos que ésta suministra. Aunque, en la práctica profesional, se suele asumir la veracidad de los testimonios y de los datos ofrecidos por los clientes o pacientes cuando describen sus comportamientos, sus estados, sus síntomas o sus problemas psicológicos, este supuesto puede ser ingenuo. Como sabemos, numerosas razones pueden dar lugar a la reserva y a la falta de cooperación, como son la búsqueda de un determinado objetivo (económico, profesional o judicial), la duda respecto a la confidencialidad de los datos, el desacuerdo con los puntos de vista y con los valores del profesional, la defensa de los propios intereses, la voluntariedad o la imposición de la evaluación o, simplemente, la inadvertencia en la realización de test y en la cumplimentación de escalas o cuestionarios. La simulación, las actitudes defensivas o de camuflaje, no obstante, no son fenómenos dicotómicos sino que suelen presentarse con niveles de intensidad variable, dependiendo de las circunstancias o de los motivos que las impulsan.

Con el propósito de analizar estas cuestiones, hemos invitado a expertos en diversos campos de la actuación profesional de los psicólogos a expresar sus puntos de vista sobre cómo el disimulo, el engaño y la mentira pueden afectar a la validez de la evaluación psicológica y cuáles son los procedimientos y estrategias utilizados para contrarrestar su efecto.

Los trabajos de Ramón Arce y Francisca Fariña, y de Verónica Godoy y Lorenzo Higuera abordan una cuestión muy polémica, como es la credibilidad del testimonio en entornos forenses. La pregunta clave es si puede la credibilidad de un testimonio ser objeto de investigación científica o si, por el contrario, ésta pertenece al ámbito de la subjetivi-

dad. En el trabajo de Arce y Fariña se presenta un procedimiento sistematizado que estos autores han desarrollado para tomar decisiones respecto a la fiabilidad y validez de las declaraciones, o la veracidad de la huella psicológica aducida o denunciada por los demandantes; mientras que el trabajo de Godoy e Higuera somete a crítica un procedimiento dudosamente válido para determinar la credibilidad de los testimonios, cual es el Análisis del Contenido Basado en Criterios (ACBC), utilizado por algunos psicólogos forenses. El artículo de Jaume Masip aborda exhaustivamente también una importante cuestión, de posible interés en el ámbito forense, como es si está fundamentada la creencia popular de que la mentira puede detectarse mejor a través de conductas no verbales que mediante el análisis de los mensajes verbales.

Otros dos artículos analizan la simulación y el engaño en el ámbito clínico. El trabajo de Mercedes Inda y cols. ofrece algunas matizaciones conceptuales y metodológicas para el estudio de las conductas de simulación; describe las condiciones clínicas en las que, con mayor frecuencia, se manifiestan dichas conductas; y presenta algunos instrumentos específicamente diseñados para su desenmascaramiento. Manuel Porcel y Rubén González, asumiendo abiertamente que la mentira y el disimulo forman parte esencial del comportamiento humano, sostienen que las conductas patológicas son mera cobertura ficticia de los problemas de la vida y que las intervenciones psicoterapéuticas, para tener éxito, deben operar bajo este supuesto.

El análisis y el manejo de la simulación y de la deseabilidad social en las evaluaciones psicológicas llevadas a cabo en los contextos organizacionales, fueron magníficamente abordados por Jesús F. Salgado; mientras que, finalmente, el disimulo y las respuestas de acomodo o deseabilidad social en el MMPI-2 son expuestos en el trabajo realizado por Héctor González Ordí e Iciar Iruarrizaga.

Agradecemos a todos los autores su participación en este monográfico, y esperamos que las contribuciones aquí presentadas serán de gran interés y utilidad práctica para los psicólogos profesionales.

## REFERENCIAS

- Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Smith, D. L. (2005). Mentirosos natos. *Mente y Cerebro*, 14, 64-69.
- Trivers, R. (2002). *Natural selection and social theory: Selected Papers of Rober Trivers (Evolution and Cognition Series)*. New York: Oxford University Press.
- Whiten, A. y Byrne, R. W. (Eds.). (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge: Cambridge University Press.



# PERITACIÓN PSICOLÓGICA DE LA CREDIBILIDAD DEL TESTIMONIO, LA HUELLA PSÍQUICA Y LA SIMULACIÓN: EL SISTEMA DE EVALUACIÓN GLOBAL (SEG)

## PSYCHOLOGICAL EVIDENCE IN COURT ON STATEMENT CREDIBILITY, PSYCHOLOGICAL INJURY AND MALINGERING: THE GLOBAL EVALUATION SYSTEM (GES)

Ramón Arce\* y Francisca Fariña\*\*

\* Departamento de Psicología Social, Universidad de Santiago de Compostela

\*\* Área de Psicología Básica, Universidad de Vigo

*El presente trabajo presenta un procedimiento de detección de la mentira y el engaño en procesos judiciales adaptado al sistema judicial español: el Sistema de Evaluación Global. El S.E.G. resulta de un compendio de diversas estrategias de análisis de la fiabilidad (CBCA y Reality Monitoring) y validez de la declaración (SVA y SRA) a la vez que de la huella psicológica y de la simulación de la misma. Además, para poder obtener unos protocolos susceptibles del presente análisis se revisan y presentan los diversos modos productivos y efectivos en la obtención de la declaración con adultos, menores, discapacitados así como la entrevista clínico-forense para la valoración del daño psíquico y de la simulación.*

*This paper presents a protocol adapted to the Spanish judicial system for the evaluation of malingering and deception in testimony, the Global Evaluation System (GES). The GES is the product of a combination of different strategies of assessment of the reliability (CBCA and Reality Monitoring) and validity (SRA and SVA) of the statements, as well as the evaluation of the psychological injury and malingering. Moreover, this procedure has been validated for different cases and witness (children, adults and handicapped), thus we review the productive and effective interview protocols for the estimation of the credibility and the psychological injury.*

**L**a formación de juicios legales y la subsecuente plasmación de los mismos en sentencias judiciales, que constituyen la pieza angular del Sistema Judicial (Sallmann y Willis, 1984), descansan en las dos dimensiones básicas formuladas por los *Modelos de Integración de la Información*: la fiabilidad y la validez (Ostrom, Werner y Saks, 1978). Sucintamente, un juicio es una evaluación de las pruebas en una dimensión. Los juicios están basados en un conjunto de creencias sobre las pruebas (v. gr., inferencias sobre los motivos del acusado, las capacidades) que son relevantes para la dimensión de evaluación, de tal modo que cada creencia tiene un peso que incide en la evaluación de la prueba para la dimensión de juicio. Este peso es conocido como el valor escalar de la creencia. Ahora bien, no toda creencia contribuye por igual a la valoración de la prueba. Dicha contribución es el resultado de la estimación de la fiabilidad y validez de la creencia. La fiabilidad en la Sala de Justicia viene definida fundamentalmente por la credibilidad de los testigos. Por su parte, la validez está determinada por la relevancia de la prueba para el jui-

cio a llevar a cabo. No obstante, la credibilidad de los testigos aporta el valor escalar más elevado en el juicio alcanzado tanto por jurados (Arce, Fariña y Real, 2000) como por jueces (Arce, Fariña, Novo y Seijo, 2001) y es la pieza clave para alcanzar juicios contrapuestos (p. e., la asignación de una alta credibilidad al testimonio de la víctima es un predictor muy robusto de culpabilidad en tanto la falta de credibilidad en el testimonio de la víctima pronostica fehacientemente un juicio de inocencia).

La estimación de la credibilidad de un testimonio viene a ser la apreciación de la exactitud que el testigo o una parte de su declaración le inspira al evaluador y le induce a creer que los hechos sucedieron tal y como declara (Mira, 1989). Para la estimación de la credibilidad se han formulado dos modelos: el modelo subjetivo o social y el modelo objetivo o científico (Vrij, 2000). Por modelo social se entiende el recurso a indicadores subjetivos no validados científicamente de credibilidad, en tanto el modelo objetivo presupone la asunción de criterios empíricos de credibilidad. En la práctica judicial diaria tanto los expertos, esto es, jueces y magistrados (Piñeiro, 2005) como los legos en derecho, o sea, jurados (Arce, Fariña y Seijo, 2003) se sirven de criterios subjetivos para la estimación de la credibilidad de los testimonios que no son eficaces en la correcta clasificación de la exacti-

*Correspondencia: Ramón Arce. Facultad de Psicología. Departamento de Psicología Social, Básica e Metodología. Universidade de Santiago de Compostela. 15706 Santiago de Compostela. España. E-mail: psarce@usc.es.*

tud del testimonio. Por ello, la Sala de Justicia ha de acudir a técnicos que valoren de un modo empírico la credibilidad de los testigos mediante técnicas productivas y científicamente validadas. Al respecto, una revisión de la literatura puso de manifiesto que los procedimientos llevados a cabo por psicólogos basados en el análisis de contenido de las declaraciones eran los más efectivos, clasificando correctamente, en contextos de simulación (o sea, con testigos simulados en casos de laboratorio), entre el 65% y el 85% de las declaraciones, en tanto que en estudios de campo (esto es, en condiciones de realidad) la eficacia llegaba al 100% (Vrij, 2000). No obstante, estos instrumentos presentan algunas limitaciones que es preciso encarar. Primera, las categorías de los diversos procedimientos propuestos para el análisis de contenido no son homogéneas inter-sistemas. Segunda, no suelen cubrir a todo tipo de testigos (generalmente se proponen para niños). Tercera, estos procedimientos no van acompañados de protocolos de actuación, a la hora de obtener la declaración, adecuados para cada contexto (menores, adultos, discapacitados). Cuarta, dichos procedimientos no analizan la totalidad de la tarea judicial (se suelen centrar en la supuesta víctima, obviando al supuesto agresor, que puede ser la víctima real). Quinta, los mismos no se orientan a la búsqueda de la huella psicológica del crimen: el trastorno de estrés post-traumático. Sexta, no incluyen metodologías adecuadas para la evaluación clínica en el contexto forense en el que debe sospecharse simulación, esto es, la mentira clínica. Séptima, los sistemas basados en el análisis de contenido de las declaraciones propuestos para el estudio de la credibilidad son técnicas semi-objetivas que precisan ajustarse para alcanzar una mayor objetividad.

Partiendo del estado de la literatura sobre la evaluación empírica de la credibilidad del testimonio y con el objetivo de afrontar las limitaciones previamente enumeradas, hemos creado (Arce y Fariña, 2002, 2005, en prensa) un protocolo psicológico forense, fruto de numerosos estudios propios y de la práctica profesional, en el que abordan la tarea en su totalidad, el *Sistema de Evaluación Global*. Para su desarrollo revisaremos inicialmente y de un modo somero los protocolos de obtención de la declaración, que constituyen requerimientos *sine qua non* se pueden aplicar los sistemas categoriales de análisis de contenido. Posteriormente, presentaremos los sistemas de análisis de la credibilidad basados en el análisis de contenido más productivos y efectivos. En tercer lugar presentaremos un protocolo de evaluación de la huella psíquica del acto delictivo y de la detección de

la simulación. Finalmente, integraremos todo ello en un procedimiento adaptado al contexto legal español que evalúa la realidad de la huella de memoria y la huella psíquica con el control de la simulación: el *Sistema de Evaluación Global*.

### LA OBTENCIÓN DE LA DECLARACIÓN

Las herramientas fundamentales para la obtención de información de los testigos son los interrogatorios y entrevistas. Es bien sabido que el éxito de dicha entrevista o interrogatorio va a depender de factores tales como la pericia del entrevistador, el grado de colaboración del entrevistado, el tiempo transcurrido desde el suceso, y, evidentemente, del tipo de entrevista (véase Memon y Bull, 1999). Los interrogatorios conforman el instrumento por excelencia de la obtención del testimonio en las actuaciones policiales y judiciales, pero la declaración obtenida mediante éstos no es productiva para la aplicación de los sistemas de análisis de la credibilidad basados en el contenido de las declaraciones. Asimismo no todos los tipos de entrevistas son válidas para estos fines. De facto, las entrevistas dirigidas o semidirigidas pueden introducir información engañosa en las narraciones de los eventos de personas sinceras (p. e., Loftus, Korf y Schooler, 1988), de modo que las distorsiones no serían fruto de la mentira, sino de la información introducida por el entrevistador. En consecuencia, se precisa de entrevistas en las que el entrevistador no mediatice la información a recuperar por el testigo. Las entrevistas en formato narrativo de discurso libre, tal como la entrevista cognitiva, se ajustan a estas demandas. Sobre éstas, Köhnken, Milne, Memon y Bull (1999), tras una revisión de la literatura, observaron que facilitaban una mayor tasa de recuperación de información, sobre todo correcta (36%), pero también incorrecta (17,5%). Ahora bien, esto no significa que la exactitud global sea mayor en un tipo de entrevista que en otro. De hecho, el porcentaje promedio de información correcta es en el formato de entrevista policial (esto es, interrogativa) del 82%, en tanto en la entrevista cognitiva se cifra en el 84%. En esta línea se han propuesto formatos de entrevista narrativa ajustados al contexto y a las capacidades del entrevistado. A continuación revisamos brevemente cada uno de ellos.

### LA ENTREVISTA COGNITIVA<sup>1</sup>

La entrevista cognitiva comprende cuatro técnicas generales de recuperación de memoria:

a) La primera técnica consiste en reconstruir mentalmente los contextos físicos y personales que existieron en el momento del crimen (o del hecho a recordar), esto es, la *reinstauración de contextos*. Esto implica que al testigo se le pida que trate de situarse mentalmente en el lugar del suceso teniendo en cuenta:

- Elementos emocionales (p.e., trata de recordar cómo te sentías).
- Elementos secuenciales (v. gr., piensa en lo que estabas haciendo en ese momento).
- Características perceptuales (vbgr., retoma la imagen de la escena del crimen y haz un dibujo de la habitación ¿A qué olía? ¿Qué oías?).

La razón que subyace a esta primera técnica es el principio de codificación específica de Tulving, esto es, la información contextual de un suceso se codifica junto con el evento y se conecta asociativamente (Tulving y Thompson, 1973). A su vez, la recuperación verbal del suceso depende del grado en que los índices ambientales de la situación a recuperar se solapan con propiedades previamente codificadas (Tulving, 1983). Esta primera técnica de la entrevista cognitiva es similar a la usada por parte de jueces y policía, la llamada reconstrucción de los hechos, con la salvedad de que la reconstrucción en la entrevista cognitiva se realiza de forma mental.

b) La segunda técnica, el *recuerdo libre*, consiste en pedirle al testigo que narre todo lo sucedido, absolutamente todo, incluyendo la información parcial e incluso detalles que considere banales para la investigación, porque éstos pueden llevar a otros, asociados entre sí en la memoria, que sean relevantes. Esta estrategia se presenta especialmente importante a la hora de combinar la información de diferentes testigos. Además, los pequeños detalles, en ciertos casos, pueden producir buenas pistas.

Con estas dos instrucciones, recuperación de la imagen mental de los hechos e informar de todos los detalles accesibles, se obtiene una primera versión de lo sucedido. Esta declaración, por tanto, es de tipo narrativo, dejando al testigo hablar sin interrupciones ni preguntas. Es preciso señalar que, en todo momento, ha de procurarse un ambiente propicio para la concentración del testigo, sin ruidos ni personas que lo distraigan, y es obvio que el entrevistador ha de ganarse la confianza del testigo para

que su declaración sea lo más sincera y productiva posible.

c) La tercera técnica, el *cambio de perspectiva*, trata de animar al testigo a que se ponga en el lugar de la víctima, o de otro testigo del suceso, incluso del sospechoso, y que informe de lo que vio o hubiera visto desde esa perspectiva, si estuviera ocupando el lugar de esa persona. Esta técnica se apoya en los estudios de Bower (1967), quien advirtió que los sujetos, al imaginarse los personajes de una historia, recordaban más detalles propios de la perspectiva del personaje con quien se habían identificado, que de otros personajes. De esta manera se obtiene una segunda versión de la entrevista, desde una perspectiva diferente.

d) El último componente es la instrucción que invita al recuerdo desde diferentes puntos de partida, el *recuerdo en orden inverso*. En otras palabras, se pretende que el individuo narre los hechos desde un orden diferente a como se desarrollaron (v. gr., desde el final al principio, desde el medio, hacia atrás) con el objetivo de recuperar pequeños detalles que puedan haberse perdido al hacer una narración de los hechos siguiendo la secuencia temporal que éstos tuvieron. Esta técnica intenta reducir los efectos que los conocimientos previos, las expectativas y los esquemas producen en el recuerdo y, además, puede ser efectiva para elicitar detalles adicionales (Memon, Cronin, Eaves y Bull, 1993). No en vano, como señalan Bower y Morrow (1990), tendemos a recordar el esquema o modelo mental que nos formamos de un evento más que el evento mismo.

La aplicación de la entrevista cognitiva no está limitada a la reproducción de un suceso de episodio único, ya que permite extender su utilización para recordar hechos que ocurren frecuentemente de una manera similar (Mantwill, Köhnken y Ascherman, 1995). La entrevista cognitiva incluye, a su vez, técnicas suplementarias tales como:

- a) Gimnasia memorística para la recuperación de la apariencia física: ¿Te recordó el intruso a alguien a quién conozcas? ¿Había algo inusual en su apariencia?
- b) Nombres: Trata de recordar la primera letra del nombre, recorriendo una a una todas las letras del alfabeto.
- c) Objetos: Describe los objetos que había dentro y fue-

<sup>1</sup> Las entrevistas han de grabarse para los posteriores análisis de contenido

- ra de la habitación ¿Parecían pesados de llevar?
- d) Conversaciones y características del habla: ¿Se utilizaron palabras extranjeras o inusuales? ¿Advertiste algún acento? ¿Tartamudeaba?
- e) Número de matrículas de automóviles: ¿Te recordaron los números o las letras de la matrícula a alguna cosa? ¿Los números eran altos o bajos?

Además de esta versión estándar de la entrevista cognitiva, Fisher y Geiselman (1992) propusieron una versión mejorada. Ésta responde a una adaptación al contexto judicial. No obstante, la efectividad y el procedimiento en términos cognitivos es el mismo. Las siguientes fases resumen la estructura general mejorada de la misma.

- Fase 1. Presentaciones y personalización de la entrevista (presentación, usar el nombre del entrevistado).
- Fase 2. Establecimiento de la comunicación (creación de una atmósfera agradable, de confianza a través de la formulación de preguntas neutras).
- Fase 3. Explicación del propósito de la entrevista.
- Fase 4. Reinstauración de contextos.
- Fase 5. Recuerdo libre.
- Fase 6. Preparación para el interrogatorio (pedirle que se concentre intensamente, que diga lo que se le viene a la mente tal como llega, sin *fabricarlo*, que puede decir *no comprendo, no sé, no recuerdo*, que active y contraste imágenes).
- Fase 7. Interrogatorio compatible con el testigo (cada testigo tiene una secuencia de memoria distinta del evento debiendo el entrevistador ajustarse a esa secuencia).
- Fase 8. Recuerdo desde diferentes perspectivas.
- Fase 9. Recuerdo en orden inverso.
- Fase 10. Resumen (realizado por el entrevistador en función de lo que el entrevistado ha informado).
- Fase 11. Cierre (desactivación emocional y de tensiones en el entrevistado).

### LA ENTREVISTA A NIÑOS TESTIGOS

Cuando los testigos son niños, el uso del procedimiento de la entrevista cognitiva merece ciertas consideraciones. Primera, uno de los hallazgos más consistentes respecto a la memoria de los niños se relaciona con el rendimiento. Sucintamente, en tareas de recuerdo libre los niños recuperan significativamente menos información que los adultos; esto es, el nivel de detalles y exactitud con el que se recuerda un suceso se incrementa con la edad (Davies, Tarrant y Flin, 1989). Segunda, cuando la tarea de recuerdo está relacionada con un contexto significati-

vo y familiar, las habilidades de memoria que exhiben los menores son superiores a cuando se trata de un contexto desconocido y vacío de significado (Bauer y Mandler, 1990). En otras palabras, en estos contextos la información que restauran no es necesariamente menos productiva y exacta. Ya que el principal objetivo de la entrevista cognitiva es aumentar la información recuperada, éste es el procedimiento más indicado para utilizar con niños (Memon y Bull, 1991). No obstante, se han de modificar las instrucciones para que el niño comprenda lo que se le está pidiendo y adaptar las demandas de la tarea a las capacidades mentales del mismo.

Con el propósito de adaptar la entrevista cognitiva a niños, el Ministerio de Interior y el Departamento de Salud Británicos (Home Office and The Department of Health, 1992; en castellano puede encontrarse desarrollado en Bull, 1997) encargaron a los profesores Diane Birch y Ray Bull la concreción de un protocolo de actuación para la obtención de la declaración de menores testigos. Sobre la base del borrador elaborado por estos profesores, junto con un foro de discusión con técnicos, se creó un protocolo de actuación específico para menores. Previamente a la entrevista en sí, recomiendan recabar información sobre el estado evolutivo del niño, nivel de lenguaje, y maduración física, social y sexual. Dicho protocolo consta de cuatro fases que debe completar el entrevistador por este orden específico: entendimiento y compenetración, recuerdo libre, interrogatorio y conclusión. La primera fase, *entendimiento y compenetración* con el menor, es de una importancia trascendental para la obtención del testimonio. De no conseguirse que el niño se relaje y se sienta cómodo, no se obtendrá un buen flujo comunicativo. Así, ha de tenerse presente que los niños generalmente están educados para no hablar con extraños. Por ello, se recomienda iniciar la tarea hablando de acontecimientos *neutros* tal como aficiones, amigos o colegio. El entrevistador también ha de mostrar un cuidado especial con los niños que puedan *sentirse culpables* o creer que han hecho *algo malo*. Por todo ello, es necesario insistir y dejar muy clara la necesidad de que digan la *verdad*. Por último, esta primera fase tiene por objeto informar al niño de lo que se espera que haga a partir de ese momento. En la segunda fase, el *recuerdo libre*, se les pide que cuenten todo lo acontecido (v. gr., ¿hay algo que te gustaría contarme? ¿sabes por qué estamos aquí?). El entrevistador puede actuar como facilitador pero nunca formular preguntas específicas. Una atención especial merecen las posibles cogniciones del niño. De hecho,



con cierta frecuencia creen que los adultos ya saben lo que ocurrió, o entienden que no deben informar sobre los hechos (como ya hemos advertido previamente, los niños generalmente están instruidos para no hablar con desconocidos). Por todo ello, es preciso insistir en que tienen que contar todo. A lo largo de la entrevista, el entrevistador debe adoptar una postura de escucha activa, esto es, controlarse de intervenir en las pausas y silencios largos.

La tercera fase, el *interrogatorio*, presenta el siguiente orden de prelación en la formulación de las cuestiones: preguntas abiertas, preguntas específicas pero no sugestivas, preguntas cerradas y preguntas profundas. Una vez el entrevistador esté totalmente seguro que el recuerdo libre ha terminado puede formular, en su caso, preguntas abiertas a fin de que el menor proporcione más información sobre algunos puntos no aclarados. Ahora bien, es preciso tener presente que las preguntas con el formato *¿por qué?* pueden originar en el niño un sentimiento de culpabilidad. En consecuencia, estas cuestiones han de formularse con ciertas cautelas. Además, tanto la reformulación de preguntas como la solicitud de repetición de una respuesta deben evitarse sobre la base de que pueden interpretarse como una crítica o una respuesta incorrecta, respectivamente. Las preguntas específicas tendrán como objeto la aclaración de algunos contenidos obtenidos previamente. Pero, ha de ejercerse un control efectivo sobre la *sugestión* inherente a este tipo de preguntas, esto es, la pregunta no llevará implícita la respuesta. Del mismo modo, las preguntas con dos alternativas de respuesta bipolares (por ejemplo, *sí vs. no*) tampoco tienen cabida en este momento de la entrevista. En todo caso, los contenidos de las preguntas estarán mediados por el nivel de desarrollo del niño. Las preguntas cerradas, por su parte, tendrán lugar si los modos anteriores no han dado los resultados apetecidos. Las preguntas con sólo dos alternativas de respuesta han de evitarse en la medida de lo posible, dado que los menores presentan una tendencia a dar la primera alternativa disponible, especialmente si es un *sí*, por lo que, de considerarse imprescindibles, se ha de proporcionar una salida de escape como *no sé* o *no recuerdo*. Finalmente, el entrevistador puede formular preguntas profundas que son aquellas que llevan implícita la respuesta. En desacuerdo con el protocolo original, desaconsejamos que se pregunte sobre la identidad del autor del delito por dos motivos. Primero, no es tarea del perito identificar al autor, sino emitir un dictamen sobre la fiabilidad de los hechos descritos. Segundo, sería muy aventurado dar

nombres, porque nuestros procedimientos de estimación de la realidad son para hechos, con lo que la transposición o cambio intencionado de personas no podría ser detectado por nuestros medios. La cuarta fase, el *cierre de la entrevista*, consiste en una recapitulación en la que se indaga, usando un lenguaje adaptado a la evolución del niño, sobre si lo obtenido en la entrevista es correcto; y una clausura, en la que se pretende establecer un nivel de angustia y tensión normalizados (al igual que al inicio de la entrevista se introducen cuestiones neutras, se le agradece la colaboración, y se le hace saber que ha respondido adecuadamente).

A este protocolo general, nosotros le hemos añadido unos procedimientos complementarios de actuación para casos específicos. Primero, cuando nos proporciona información sobre conceptos que el menor aún no ha operativizado adecuadamente, tal como la estimación de tiempos, longitudes u otros, acudimos a comparaciones con otros referentes bien establecidos. Así, si queremos conocer la extensión temporal de un determinado evento lo podemos comparar con un recreo (p. e., *¿fue más corto, más largo o igual que un recreo?* Si la respuesta fuera más corto, ésta nos indicaría que muy probablemente duró menos de media hora). Segundo, procedemos con otros lenguajes complementarios para aquellos casos en que el nivel de desarrollo lingüístico del menor así lo aconseja; esto es, en ocasiones el menor tiene las imágenes de los hechos, pero no las destrezas lingüísticas suficientes para transformar algunas imágenes mentales en palabras, o la falta de habilidades es tal que la declaración es muy corta y, por tanto, no es productiva para nuestros propósitos. En estos casos, el recurso a otros procedimientos de comunicación con el menor, no sólo son adecuados, sino también convenientes. Los sistemas de comunicación a los que recurrimos son los trazos de dibujos, la representación de los hechos o el uso de utensilios, tales como bolígrafos, con los que se les pide que ejecuten una determinada imagen mental que poseen. Obviamente, estos lenguajes son complementarios y no deben usarse en aislado, porque pueden inducir fácilmente a un error de interpretación. Pero, cuando describen o intentan describir hechos que no comprenden y su vocabulario no está formado pueden dibujarlo, señalarlo en un modelo o en sí mismos, hacer una representación de los mismos o reproducir la acción con objetos. Ha de tenerse presente que entre los objetos no es aconsejable que se usen *muñecos anatómicamente correctos*, ya que conllevan una alta probabilidad de error en las interpretaciones del perito (Dammeyer, 1998), al tiempo

que la memoria elicitada vía estos muñecos también incrementa el número de respuestas incorrectas entre los más niños (Goodman y otros, 1997). En todo caso, recordamos que la información obtenida por estos medios ha de ser únicamente aclaratoria o de investigación, y nunca por sí sola prueba de acusación.

¿Cómo saber qué procedimiento seguir para recurrir a la entrevista cognitiva o al protocolo de menores? La diferencia básica de ambos procedimientos está en las habilidades cognitivas que se le exigen al menor. La entrevista cognitiva requiere de capacidad de empatía para, por ejemplo, el cambio de perspectiva. Así, si ésta se adquiere de forma gradual a partir de los 8-9 años (Vrij y Winkel, 1996), es aconsejable recurrir al protocolo de niños con edades anteriores a éstas. En todo caso, los menores de 7 años tienen dificultades para seguir las técnicas comprendidas en la entrevista cognitiva (Memon, Cronin, Eaves y Bull, 1996). Si bien, la entrevista cognitiva puede ser efectiva en menores de edades superiores a 7 años, se corren determinados riesgos tal como que las respuestas estén mediatizadas por las características de la demanda (Memon, Wark, Bull, y Köhnken, 1997). En los otros casos, es recomendable iniciar con el protocolo de entrevista cognitiva y de observarse falta de destrezas cognitivas en el entrevistado cambiar al protocolo de menores. Nosotros hemos comprobado que este sistema es factible en la práctica, no dando lugar a distorsiones.

### LA ENTREVISTA A PERSONAS DISCAPACITADAS

La literatura científica sobre la entrevista a personas discapacitadas es muy poco prolífica, siendo necesaria más investigación sobre las técnicas concretas a aplicar (Bull, 1995). De hecho, se ha encontrado que mediante la entrevista cognitiva recuperan en torno a un 32% más de información correcta, pero, al mismo tiempo, aumenta significativamente el número de confabulaciones. En todo caso, deben considerarse aún más los problemas de las preguntas sugestivas, de las preguntas cerradas y profundas. En esta línea, Cahill y otros (1988, citado en Bull 1995) han confeccionado un listado de aspectos que se han de evitar:

- a) La aquiescencia del testigo con preguntas sugestivas, de modo que la respuesta sea la solicitada.
- b) Una presión indebida que conduzca al testigo a confabular (p. e., sentirse parte de un evento del que no ha sido testigo).
- c) Preguntar repetidamente sobre un punto particular, provocando que los testigos establezcan conjeturas

o se desvíen de su respuesta inicial (las preguntas repetidas les conducen a asumir que la respuesta no era correcta).

- d) Que el entrevistador no se precipite en etiquetar de ambiguo o pobre el lenguaje usado por parte de estos testigos.
- e) Que el entrevistador ofrezca descripciones a los testigos que tienen dificultad en encontrar sus propias palabras (v. gr., si la chaqueta no era oscura o clara, entonces ¿dirías que era una especie de color marrón?).
- f) Que el entrevistador proporcione alternativas de respuesta cerradas a los testigos (vbgr., ¿llevaba un revolver o una escopeta?).
- g) Cuando el testigo utilice una muletilla tal como *¿no sabes?*, el entrevistador debe proceder de modo que no se quede sin información (una posibilidad es advertir directamente al testigo que el entrevistador no sabe, que se explique).
- h) No ignorar un fragmento previo de información proveniente del testigo que no se ajuste con la asunción que el entrevistador tiene sobre lo que ha ocurrido.
- i) Que el entrevistador no comprenda todo lo que el testigo narra.
- j) Que el entrevistador no compruebe, usando los medios apropiados, que ha comprendido al testigo.

Ante este estado de la literatura y respondiendo a las demandas que se han formulado desde las Salas de Justicia, Arce, Novo y Alfaro (2000) crearon unas pautas y un protocolo de actuación en estos casos. Previamente a proceder a realizar una evaluación del testimonio de una persona supuestamente discapacitada, es preciso determinar su capacidad como testigo. No en vano nuestro sistema jurídico prevé, primeramente, la obligación general de declarar: La LECrim, en el Artículo 410, fija que toda persona que viva en el territorio español, nacional o extranjero, está obligado a concurrir al llamamiento judicial para declarar. Pero, asimismo, también establece una serie de exenciones a la obligación a declarar entre los que se incluyen los incapacitados física o moralmente (V. Artículo 417, párr. 3). El procedimiento se inicia con la cumplimentación de las escalas de Wechsler correspondiente, generalmente el WAIS. Éstas nos proporcionan información muy fiable sobre la capacidad intelectual del individuo, y, además, son un indicador robusto de lesión cerebral. En concreto y en el WAIS, nos alertan de lesión posiblemente relevante aquellos resultados con una diferencia significativa entre las subescalas verbal y manipulativa (unos autores apun-

tan >10 puntos en tanto otros los elevan a 15) (véase Wechsler, 1976 para una revisión) y, en el caso que nos atañe, debe ser mayor el coeficiente manipulativo que el verbal (posible organicidad en el hemisferio izquierdo). Asimismo, las entrevistas y antecedentes del sujeto aportarán información importante para definir si es o no factible una lesión a la vez que el grado de minusvalía. Sin embargo, no toda lesión cerebral incapacita a los sujetos para dar información, incluidos aquellos casos que puedan ser incapacitados como testigos en la Sala de Justicia. Así pues, se procede, una vez detectada una posible lesión cerebral o deterioro mental, a identificar las áreas sobre todo cognitivas afectadas por la lesión o deterioro. Las herramientas aconsejables son el Test Barcelona (Peña-Casanova, 1990) y el ERFC (Gil, 1999) por su fiabilidad y aunar en un único instrumento la evaluación de las áreas que permiten ajustar al testimonio sobre los hechos judiciales las limitaciones que tiene el testigo. En una primera aproximación, las lesiones operativas se clasifican en afasias, alexias y acalculias. Obviamente, las lesiones suelen centrarse en alguna de ellas. No obstante, cada agrupación se subdivide en áreas. Tras la evaluación global deberá realizarse una evaluación de la adecuación del testimonio (esto es, concreción de las áreas accesibles, las deficitarias y de los modos de obtención del testimonio), y bajo qué condiciones se podría obtener un testimonio fiable (p. e., un testimonio ante un forense experto en este tipo de evaluaciones puede ser válido, pero no así ante un juez, policía o abogados). Así, por ejemplo, una persona con anomia y praxis orofonatoria puede evidenciar dificultades en identificar comportamientos verbalmente e intercambiar palabras (v. gr., vejiga por vagina), lo cual no indica que carezca de fiabilidad en el testimonio aportado. También es frecuente encontrar testigos que presentan anomalías con claras implicaciones para la obtención de la declaración en la tarea *memoria verbal de textos*. En concreto, puede que no sean capaces de recuperar directamente el evento, pero sí de responder a un interrogatorio sobre el mismo, esto es, carecen de memoria episódica pero no de memoria de los eventos si es guiada por preguntas. En consecuencia, pueden responder perfectamente a un interrogatorio y no tienen capacidad para crear un testimonio falso congruente al adolecer de memoria episódica (hemos observado este cuadro entre consumidores habituales de drogas). En el más adverso de los casos, o sea, cuando la deficiencia es severa, los sujetos pueden aportar igualmente información ya que las personas con daño neurológico, siem-

pre que tengan la visión intacta, pueden almacenar y recuperar información visual (Freed y otros, 1989; Hart y O'Shanick, 1993; Winograd, Smith y Simon, 1982). De facto, sólo sería necesaria la reinstauración del contexto (téngase presente que muchos de ellos padecen de entecimiento con lo que puede demorarse y, por tanto, se debe ser muy paciente a fin de no truncarlos en medio de la búsqueda). Bajo esta contingencia, no es posible una mentira efectiva, y lo trascendente es buscar un sistema de comunicación que vendrá definido por el análisis neurológico, siendo el más accesible la representación de los hechos. Este procedimiento se ha mostrado productivo con muy diferentes tipos de deficiencias y ha sido validado judicialmente en varios casos.

### LA ENTREVISTA CLÍNICO-FORENSE

Una última fuente de obtención de información con implicaciones para la fiabilidad de un testimonio procede del ámbito clínico. Los instrumentos usuales de medida clínica están desarrollados sobre la base de que estamos ante un paciente. Por tanto, no tiene interés el estudio de la simulación. Por ello, las entrevistas estructuradas o semi-estructuradas, al igual que los listados de síntomas e instrumentos de medida psicométricos, no cumplen con el propósito de controlar la simulación de un trastorno mental al propiciar información que la facilita. Por ejemplo, ante la pregunta *¿tiene usted dolores de cabeza?* (1ª pregunta del SCL-90-R, Derogatis, 2002), el sujeto simulador tiene ante sí un efecto facilitador de respuesta de simulación. Este tipo de preguntas proporcionan al sujeto *un camino conducente* para la selección de los síntomas asociados a una determinada enfermedad psíquica, con lo que ya sólo sería suficiente que tenga la habilidad precisa para discriminar entre ítem pertenecientes a una patología u otra. Al respecto, los datos muestran que no hay referencias basadas en la entrevista clínica tradicional que lleguen al diagnóstico de simulación (v. gr., Rogers, 1997), a la vez que los sujetos son capaces de simular efectivamente una enfermedad y discriminarla de otras (p. e., Arce, Carballal, Fariña y Seijo, 2004; Arce, Fariña y Pampillón, 2002). Si bien algunos instrumentos psicométricos cuentan con escalas del control de *validez* de los datos del registro, no son prueba suficiente para establecer fehacientemente simulación porque: a) el diagnóstico de simulación es compatible con la formulación de otras hipótesis alternativas (p. e., Graham, 1992; Roig-Fusté, 1993), b) no clasifica correctamente a todos los simuladores (v. gr., Bagby, Buis y Nicholson, 1995), y c) no proporcionan diagnósticos sino impresio-

nes diagnósticas. En consecuencia, la toma de decisiones fundada única y exclusivamente en estos instrumentos deja la puerta abierta a la comisión sistemática de dos tipos de errores: falsos positivos (catalogar a enfermos reales como simuladores) y errores de omisión (no detectar como simuladores a sujetos que realmente están simulando). Todo ello ha llevado a que se proponga, a fin de minimizar estas fuentes de error, la adopción de una estrategia de evaluación multimétodo (p. e., Rogers, 1997). En este contexto, es donde tiene cabida una entrevista de orden clínico que permita un diagnóstico y sirva de contraste a los datos obtenidos por otros métodos. Así, hemos concretado lo que denominamos *Entrevista Clínico-Forense*. La entrevista, llevada a cabo por un entrevistador entrenado y con conocimientos de psicopatología, consiste en pedir a los sujetos que relaten en un discurso narrativo libre los síntomas, conductas y pensamientos que tienen en el momento presente en comparación con el estado anterior al delito (esto es, EE-AG en el eje V del DSM-IV-TR). Si los sujetos no responden de motu propio, les será requerido por medio de preguntas abiertas, de acuerdo con el eje V del DSM-IV-TR (American Psychiatric Association, 2002), que informen igualmente sobre sus relaciones familiares (EEGAR); relaciones sociales (EEASL) y relaciones laborales (EE-ASL). Con este procedimiento, se solicita a los sujetos la ejecución de una tarea de conocimiento de los síntomas que padecen en tanto con las entrevistas estructuradas, semi-estructuradas, listados de síntomas e instrumentos psicométricos desempeñan una tarea de reconocimiento de síntomas. Es por ello que la entrevista no es en formato de interrogatorio, sino *no directiva* y orientada a la reinstauración de contextos. O sea, seguimos el procedimiento de entrevista abierta y en formato de discurso libre seguida de una reinstauración de contextos. Este procedimiento de entrevista se mostró fiable, válido y productivo en la detección de simulación de un trastorno de estrés postraumático ligado a una falsa agresión sexual e intimidación (Arce, Fariña y Freire, 2002), violencia de género (Arce y otros, 2004), accidente de tráfico (Arce, Fariña, Carballal y Novo, 2006) y de un trastorno mental no imputable (Arce, Fariña y Pampillón, 2002).

La obtención de esta información clínica debe ser grabada y procederse al análisis de contenido de la misma. Las categorías de análisis son los síntomas descritos en el DSM-IV-TR. Así, creamos un sistema categorial mutuamente excluyente, fiable y válido, en lo que Weick (1985) ha denominado sistemas de categorías metódicas. Una vez construidas las hojas de registro se marcan

los diferentes síntomas detectados. Si bien la gran mayoría de los síntomas, incluidos los más adversos, pueden ser informados directamente por los sujetos (Lewis y Sarni, 1993), algunos sólo pueden observarse. En consecuencia, la detección de las categorías responde a dos métodos complementarios: expresión directa del sujeto e inferencias de los codificadores tras analizar los protocolos. Por ejemplo, el deterioro de memoria puede ser manifestado directamente por el sujeto o ser inferido por el codificador tras la entrevista.

### EN BUSCA DE LA VERDAD: ANÁLISIS DE CONTENIDO DE LA DECLARACIÓN

La revisión y puesta en común de la literatura sobre análisis de contenido de las declaraciones nos han llevado en el *Sistema de Evaluación Global* a fraccionar la estimación de la credibilidad de las declaraciones de los testigos en función de dos parámetros: la validez y la fiabilidad. La validez sirve para establecer la admisibilidad de la prueba para el análisis de contenido en tanto la fiabilidad se relaciona con los indicios de realidad contenidos en la declaración.

### ANÁLISIS DE LA VALIDEZ DE LA DECLARACIÓN

El estudio de la validez como prueba de la declaración se aborda a partir del procedimiento completo (v. gr., declaraciones ante la policía o juez, otros testimonios, otras pruebas insertas en el procedimiento) y de las grabaciones de las declaraciones prestadas ante los peritos. Dos son los sistemas que dan entrada a un análisis sistemático de la validez de las declaraciones: El SRA y el SVA.

El SRA, Statement Reality Analysis (Undeutsch, 1967, 1988), aborda el estudio de la validez de la declaración a través de las siguientes categorías:

- a) Criterios negativos o de control:
  - Carencia de consistencia interna (contradicciones).
  - Carencia de consistencia con las leyes de la naturaleza o científicas.
  - Carencia de consistencia externa (discrepancia con otros hechos incontrovertibles).
- b) Criterios derivados de las secuencias de declaraciones:
  - Carencia de persistencia (estabilidad en el tiempo y contextos).
  - Declaración inconsistente con la anterior.

Por su parte, el SVA, Statement Validity Analysis (p. e., Steller, 1989), recoge las siguientes categorías de evaluación:



- a) Características psicológicas:
  - Adecuación del lenguaje y conocimientos.
  - Adecuación del afecto.
  - Susceptibilidad a la sugestión.
- b) Características de la entrevista:
  - Preguntas coercitivas, sugestivas o dirigidas.
  - Adecuación global de la entrevista.
- c) Motivación:
  - Motivos del informe.
  - Contexto del informe o declaración original.
  - Presiones para presentar un informe falso.
- d) Cuestiones de la investigación:
  - Consistencia con las leyes de la naturaleza.
  - Consistencia con otras declaraciones.
  - Consistencia con otras pruebas.

Como sistema de evaluación global de la declaración proponen el mejor ajuste a una de las siguientes categorías: *creíble, probablemente creíble, indeterminado, probablemente increíble o increíble*.

### LA FIABILIDAD DE LA DECLARACIÓN

El estudio de la fiabilidad de las declaraciones, esto es, la búsqueda de criterios de realidad en los contenidos de las declaraciones, que se lleva a cabo a partir de las grabaciones de las declaraciones prestadas ante los peritos, ha conformado la gran aportación de la Psicología Forense a la valoración de la prueba. Tres son los sistemas categoriales, basados en el análisis de contenido, propuestos que se han mostrado productivos y efectivos en la estimación de la fiabilidad de la prueba: el Reality Monitoring, el SRA y el CBCA. El Reality Monitoring, en la propuesta seminal (Johnson y Raye, 1981), establece que las declaraciones verdaderas contienen más atributos contextuales (espacio-temporales) y sensoriales (sonidos, olores, etc.) en tanto las fabricadas incluyen más operaciones cognitivas, esto es, información idiosincrásica (por ejemplo, yo pensé, recuerdo ver, me sentía nervioso). Spörer (1997) amplió a ocho la lista de criterios: claridad (claridad, viveza en vez de vaguedad), información perceptual (información sensorial tal como sonidos, gustos o detalles visuales), información espacial (lugares, ubicaciones), información temporal (ubicación del evento en el tiempo, descripción de secuencias de eventos), afecto (expresión de emociones y sentimientos sentidos durante el evento), reconstrucción de la historia (plausibilidad de reconstrucción del evento tras la información dada), realismo (plausibilidad, realismo y sentido de la historia) y operaciones cognitivas (descripciones de inferencias hechas por otros durante el evento). Los

siete primeros se vinculan a veracidad y el octavo a falsedad, resultando más efectiva esta nueva recategorización. El contraste de los resultados de la declaración con las prescripciones del modelo es el procedimiento habitual de validación del origen de los atributos de memoria, pero ésta también puede llevarse a cabo a través de un proceso de razonamiento que implica el análisis de las características cualitativas del trazo, las características de los trazos relacionados, y las suposiciones mnésicas.

El SRA (Undeutsch, 1967, 1988), también contiene categorías para la evaluación de la realidad de la declaración. Éstas son:

- a) Criterios generales, fundamentales:
  - Anclaje, fijación espacio-temporal (concreción de la acción en un espacio y tiempo).
  - Concreción (claridad, viveza).
  - Riqueza de detalles (gran cantidad de detalles en la narración).
  - Originalidad de las narraciones (frente a estereotipos o clichés).
  - Consistencia interna (coherencia lógica y psicológica).
  - Mención de detalles específicos de un tipo concreto de agresión sexual.
- b) Manifestaciones especiales de los criterios anteriores:
  - Referencia a detalles que exceden la capacidad del testigo (que van más allá de su imaginación o capacidad de comprensión).
  - Referencia a experiencias subjetivas (sentimientos, emociones, pensamientos, miedos).
  - Mención a imprevistos o complicaciones inesperadas.
  - Correcciones espontáneas, especificaciones y complementaciones durante la declaración.
  - Autodesaprobación (declaración en contra de su interés).

Con todos estos criterios de decisión se procede a una evaluación conjunta, en la que los dos factores *criterios generales* y *manifestaciones especiales de los criterios generales* ponderan positivamente hacia la veracidad, esto es, la presencia de estos criterios indica que la declaración es verdadera, pero su ausencia no implica que sea falsa. Por su parte, la presencia de los criterios de validez los *criterios de control* y los *criterios derivados de las secuencias de declaraciones* restarían valor de verdad a la declaración. En todo caso, debe tenerse presente que cada criterio tiene un peso limitado en la

determinación categórica (verdadera vs. falsa) o del grado en que una declaración representa algo vivido por el testigo. Además, prescribe el seguimiento de cuatro máximas en la determinación de si la narración describe un evento real o no:

- La intensidad o grado de las manifestaciones en los diferentes criterios.
- El número de detalles de la narración que se relacionan con un criterio (o más).
- Las capacidades del declarante para informar (edad, inteligencia, sugestión).
- Las características del evento narrativo (p. e., complejidad, relevancia).

En 1994, Steller y Köhnken propusieron, a partir de las aproximaciones anteriores, un sistema integrado de categorías que tiene por objeto la evaluación de las declaraciones de menores víctimas de abusos sexuales. El CBCA consta de cinco categorías principales con 19 criterios a evaluar, que son:

a) Características generales:

- Estructura lógica (coherencia y consistencia interna).
- Elaboración inestructurada (presentación desorganizada).
- Cantidad de detalles (abundancia de detalles o hechos distintos).

b) Contenidos específicos:

- Engranaje contextual (ubicación de la narración en un espacio y tiempo).
- Descripción de interacciones (cadena de acciones entre el testigo y otros actores).
- Reproducción de conversaciones (réplica de conversaciones).
- Complicaciones inesperadas durante el incidente (por ejemplo, interrupción imprevista).

c) Peculiaridades del contenido:

- Detalles inusuales (detalles con baja probabilidad de ocurrencia).
- Detalles superfluos (detalles irrelevantes que no contribuyen significativamente a los hechos).
- Incomprensión de detalles relatados con precisión (explicitación de detalles que el menor no comprende pero realmente sí tienen sentido).
- Asociaciones externas relacionadas (inclusión de información externa a los hechos en sí pero relacionada con ellos, tal como en una agresión sexual recordar conversaciones anteriores sobre este tema).
- Relatos del estado mental subjetivo (referencias a

sentimientos, emociones o cogniciones propias).

- Atribución al estado mental del autor del delito (referencias al estado mental del agresor y atribución de motivos).

d) Contenidos referentes a la motivación:

- Correcciones espontáneas (correcciones espontáneas o mejoras de la declaración).
- Admisión de falta de memoria (reconocimiento de lagunas de memoria).
- Plantear dudas sobre el propio testimonio.
- Auto-desaprobación (actitud crítica sobre su propia conducta).
- Perdón al autor del delito (la declaración de la víctima favorece al acusado, o evitación de más acusaciones).

e) Elementos específicos de la agresión:

- Detalles característicos de la ofensa (descripciones que contradicen las creencias habituales sobre el delito).

Estos criterios de contenido pueden analizarse como presentes o ausentes, o puntuarse en cuanto a fuerza o grado en que aparecen en la declaración. En cualquier caso, éstos, si se manifiestan, se interpretarán en el sentido de que la declaración es verdadera en tanto que de su ausencia no puede desprenderse que sea falsa. Sobre el punto de corte para discriminar entre declaraciones propias o no de realidad, Steller (1989) encontró que las declaraciones verdaderas contenían, al menos, 7 criterios de veracidad.

### EN BUSCA DE LA HUELLA PSÍQUICA DEL DELITO: LA DETECCIÓN DE LA SIMULACIÓN EN LA EVALUACIÓN CLÍNICA

Se entiende por victimación el hecho de haber sido objeto de un delito pero también la amalgama de consecuencias de este acto. Éstas pueden ser de muy diversos tipos: físicas, económicas, sociales o psíquicas. La huella psíquica, al igual que la huella de memoria, puede en una evaluación pericial convertirse en prueba de cargo. Pero, en un contexto como el que nos ocupa, el médico-legal, no es suficiente con diagnosticar un trastorno o trastornos, sino que ha de sospecharse simulación (American Psychiatric Association, 2002). Para este doble objetivo, el diagnóstico clínico y control de la simulación, no es efectiva la evaluación clínica ordinaria. De hecho, la evaluación clínica tradicional nunca ha informado de simulación (i.e., Rogers, 1997). Para la medida de la huella psíquica y el control de la simulación (hipótesis a contrastar en la medida de la huella psíquica de un deli-

to), Arce, Fariña y Pampillón (2002) han creado y validado un protocolo de actuación en función de las respuestas y estrategias seguidas por los simuladores. Éste parte de la distinción operativa entre criterios positivos, que validan el protocolo, y negativos, que invalidan o mitigan la validez del protocolo, de la simulación. Los criterios positivos serían aquellos que no se detectan en los protocolos de los simuladores y éstos fueron *la no evitación de respuestas* y *la deseabilidad social*. En concreto, aquellos sujetos que sean evaluados por las escalas de control del MMPI en el sentido de que se abstienen significativamente de dar respuestas (Escala de Interrogantes) y que tiendan a dar respuestas de deseabilidad social, no siguen las estrategias típicas de simulación, lo que debe interpretarse más como indicio de veracidad del protocolo que un intento de simulación. Es preciso tener en mente que la ausencia de colaboración con la evaluación (no respuestas) se había propuesto como un indicador fiable de simulación (v. gr., Rogers, 1992; Lewis y Saarni, 1993; Bagby y otros, 1997), pero esta contingencia nunca fue observada entre los simuladores en una evaluación forense. Los criterios negativos, esto es, observados en los protocolos de los simuladores, fueron: 1) los sistemas de medición, MMPI, entrevista u otros, no detectan, en protocolos válidos, enfermedad mental (en otras palabras, si los instrumentos de medida no detectan enfermedad mental alguna no se puede sostener la misma en el contexto legal); 2) las escalas de control de validez del MMPI y sus combinaciones detectan simulación; 3) la detección de alguna estrategia de simulación en la entrevista; y 4) la ausencia de concordancia inter-medidas. El primer criterio es eliminatorio, o sea, si la enfermedad mental no es medible, no se puede sostener una huella psíquica en la victimación. Los restantes, en sí mismos, no son determinantes, con lo que se requerirá, cuando menos, la complementación de dos criterios y el estudio de las hipótesis alternativas para concluir en relación a la simulación de la huella psíquica. Para estos últimos indicios de no validez es para los que formulamos el concepto de *invalidéz convergente* que requiere de, al menos, dos indicios totalmente independientes de invalidéz para desestimar un protocolo como inválido. En función de estos criterios se formuló la siguiente propuesta de protocolo de actuación:

- a) Recurso a sistemas de medida complementarios y concordantes, que presupongan la ejecución de tareas distintas y que contengan sistemas de control de la validez. Así, se propone someter al sujeto a una evaluación psicométrica, que implica una tarea

de reconocimiento de síntomas, y a una medida mediante una tarea de conocimiento, la *Entrevista Clínico-Forense*. En relación al instrumento psicométrico a utilizar, el MMPI es el instrumento de referencia para la evaluación de daño psíquico en la práctica forense (Butcher y Miller, 1999), pero requiere de un alto nivel de comprensión por parte del evaluado. Cuando éste presenta dificultades para cumplimentar el MMPI, el checklist SCL-90-R permite superar ese inconveniente al tiempo que cuenta con medidas de control de la validez del protocolo. Por ello, recomendamos este instrumento como sustituto del MMPI o, en caso de duda, como complementario. La primera medida a tomar ha de ser a través de la entrevista para controlar el efecto del aprendizaje de la tarea psicométrica en la tarea de conocimiento. En lo referente a la concordancia inter-medidas debe tenerse presente que ésta no va a ser total. No se nos debe escapar que incluso las medidas test-retest distan mucho de la perfección.

- b) Análisis de la consistencia interna de las medidas: escalas de control de los instrumentos psicométricos, y, en la entrevista, análisis de contenido en busca de las estrategias habituales de simulación. Las escalas de control de validez del MMPI-2 (Hathaway y Mckinley, 1999) con implicaciones para el estudio de las simulaciones según este protocolo son las escalas originales de validez (escalas de no respuestas, L, F y K), los indicadores adicionales de validez del protocolo (F posterior, TRIN, VRIN), así como los índices que se han mostrado efectivos en la detección de la simulación de trastornos, el *índice F-K*, y el *perfil en V invertida* (Duckworth y Anderson, 1995). Si la evaluación psicométrica se obtiene a través del SCL-90-R (Derogatis, 2002), las escalas de validez serían la PST, PSI, GSI y PSDI. Por su parte, las entrevistas se someten a un análisis de contenido tomando como categorías las estrategias que siguen los simuladores en las entrevistas: evitación de respuestas, síntomas raros, combinación de síntomas, síntomas obvios, consistencia de síntomas, síntomas improbables, agrupación indiscriminada de síntomas y severidad de síntomas.
- c) Es aconsejable que dos evaluadores, por separado, lleven a cabo la evaluación, con lo que se podrá contrastar la consistencia inter-evaluador. Sucintamente, esta salvaguarda sirve para controlar posibles sesgos de medida e interpretación en el evaluador.

- d) Estudio de la fiabilidad de la evaluación: consistencia interna, inter-medidas, inter-contextos (antecedentes, pruebas documentales, etc.), e inter-evaluadores (Wicker, 1975).
- e) Control de falsos positivos, esto es, enfermos reales, a través de un estudio de los antecedentes e historia general del sujeto, de las hipótesis alternativas en cada indicador de no validez (véase Roig Fusté, 1993, Graham, 1992), y del cumplimiento de los criterios del Modelo de Decisión Clínica para el establecimiento de la simulación (Cunnien, 1997).
- f) Anamnesis o estudio de los antecedentes. Con éste se pretende, en su caso, reforzar la evaluación con los antecedentes del sujeto, datos del entorno, un estudio de su comportamiento, el recabamiento de pruebas documentales, otros testimonios, etc.
- g) Estudio psicológico de la huella psíquica. Las medidas clínicas proporcionan datos respecto a lo que legalmente se denomina *evaluación biológica*, pero ésta requiere, también según la demanda legal, de lo que llaman *evaluación psicológica* en la que se clarifique la relación entre la huella psíquica medida y la huella psíquica esperada para ese caso.
- h) Finalmente, también se puede poner a prueba la validez discriminante. En otras palabras, realizar una medida no relacionada con el caso, tal como de valores o personalidad (16-PF, SIV), esperando que no presente relación alguna con la evaluación objetivo y se descarte un intento de manipulación de la imagen, tanto en positivo como en negativo.

La impresión resultante sobre la simulación se ha de ajustar a las siguientes categorías: *probablemente simulador* o *probablemente no simulador*. Es importante dejar a un lado la intención de establecer la certeza (por ejemplo, situar la impresión en una escala de varios puntos) porque crea confusión a los decisores (p. e., sentencia del TS de 29 de octubre de 1981, RA 3902), al tiempo que es importante utilizar términos probabilísticos puesto que la evaluación psicológica está sujeta a error.

### EL SISTEMA DE EVALUACIÓN GLOBAL

El *Sistema de Evaluación Global* (SEG) se estructura en torno a 9 tareas que se describen y justifican brevemente a continuación: obtención de la declaración, repetición de la declaración, contraste de las declaraciones recabadas a lo largo del procedimiento judicial, análisis del contenido de las declaraciones, análisis de la fiabilidad de las medidas, medida de las secuelas clínicas del hecho traumático, evaluación de la declaración de los ac-

tores implicados, análisis de la personalidad y capacidades de los actores implicados, y, finalmente, implicaciones para la presentación del informe. Las tareas a ejecutar están mediadas por el caso a evaluar (p. e., si en un caso no es posible la evaluación del acusado, esa fase no se lleva a cabo). Las fases en la versión más completa del sistema son:

- a) *Obtención de la declaración (la huella de memoria)*. Para que el procedimiento de evaluación psicológico forense de las declaraciones y la huella psíquica sea productivo, fiable y válido requiere de unos instrumentos de obtención de la declaración y medida del estado clínico que permitan los subsecuentes análisis de las mismas. Por ello, las declaraciones han de obtenerse, según se trate de adultos, menores o discapacitados, por los siguientes procedimientos: *Entrevista Cognitiva Mejorada* (Fisher y Geiselman, 1992), *Memorandum de Buenas Prácticas* (Bull, 1997), o la *Entrevista Forense a Discapacitados* (Arce, Novo y Alfaro, 2000). Por su parte, el daño o huella psíquica en *tarea de conocimiento* se obtiene a través de la *Entrevista Clínico-Forense* (Arce y Fariña, 2001; Arce, Fariña y Freire, 2002; Arce, Pampillón y Fariña, 2002).
- b) *Repetición de la obtención de la declaración*. Los métodos propuestos se basan en una única medida de la huella de memoria. Ahora bien, con una única medida perdemos la posibilidad de análisis de la consistencia de la declaración en el tiempo (posteriormente se discute sobre la validez de las otras obtenidas a lo largo del proceso judicial). En otras palabras, no se da entrada a uno de los modos de contraste de la validez de la información: la consistencia temporal o intra-testigo (p. ej., Wicker, 1975; Schum, 1977). Asimismo, la doctrina legal ha perfilado la fiabilidad de un testimonio en función de los criterios de oportunidad (oportunidad para observar, etc.), sesgo (control de los posibles intereses), consistencia temporal, plausibilidad, consistencia inter-testigos y crédito (Schum, 1977). También nuestra jurisprudencia (p. e., Sentencia del TS de 29 de abril de 1997) establece, cuando el testimonio de la víctima sea la única o la prueba central de cargo, que el testimonio debe reunir las 3 siguientes características: ausencia de incredulidad subjetiva, alguna corroboración periférica de carácter objetivo y persistencia en el tiempo sin ambigüedades ni contradicciones. En esta línea, ya se han dictado sentencias que han anulado el valor de prueba del



análisis de contenido (CBCA y SVA) de las declaraciones basados en una única declaración (v. gr., AP de Pontevedra, Sección 6ª, de 21 de enero de 2004). En suma, tanto la metodología científica como la doctrina legal y la jurisprudencia nos demandan más de una declaración para el estudio de la consistencia temporal. Al respecto, se ha encontrado que la repetición de la obtención de la declaración no tiene porqué contaminar los datos procedentes de una entrevista no viciada externamente (v. gr., Campos y Alonso-Quecuty, 1999), tal y como ocurre en los protocolos de obtención de la declaración anteriormente mencionados. En consecuencia, en la primera medida no se ha de proceder de ningún modo con un interrogatorio, esto es, sólo se ha de acudir a la reinstauración de contextos, recuerdo libre, cambio de perspectiva y recuerdo en orden inverso. El interrogatorio, de ser necesario, se deja para la segunda medida a fin de no contaminar la memoria de eventos con el interrogatorio. De una segunda medida se obtiene un análisis de la consistencia que, de acuerdo con la hipótesis Undeutsch (1967, pág. 125), debe entenderse en función de la centralidad/periferia del material que entra en contradicción. Así, señala que sólo es relevante la contradicción si afecta a detalles centrales para la acción de juicio. La inconsistencia en la información periférica o la omisión de cierta información sólo es importante si es trascendente para la construcción de un evento verdadero. Nosotros, para dar cabida a las interferencias (teoría de la interferencia del olvido), a la entrada de nueva información (hipótesis constructiva del olvido) y a la curva del olvido, estimamos el tiempo a transcurrir entre entrevista y entrevista en superior a 1 semana (pero no mucho más allá). Los axiomas que asumimos al respecto son tres. Primero, como quiera que el acto delictivo constituye un *evento vital estresante* el *efecto del desuso* será menor (referido al testimonio tanto del denunciante como del denunciado, y contiguo a los hechos). Segundo, una *teoría de racionalidad* por parte del mentiroso que se plasma en que la mentira es planificada, aprendida y, por extensión, consistente en el tiempo con lo que no estará mediada por interferencias e información post-suceso (*hipótesis constructiva*). Para esto, es imprescindible obtener la primera declaración en formato de discurso libre sin ningún interrogatorio porque a través de éste entraría información post-

suceso que el sujeto acomodaría a la nueva reconstrucción. El interrogatorio sólo procedería tras la obtención en discurso libre de la segunda declaración. Tercero, el sujeto que dice verdad narra imágenes con lo que la descripción de los hechos aún siendo muy semejante, será de construcción distinta al no responder a esquemas episódicos. En suma y en formato de recuerdo libre, la declaración verdadera será menos consistente y aunque el evento sea el mismo, la narración será significativamente distinta tanto en su recuperación como en el contenido (omisiones, elicitación de eventos distintos a los hechos pero relacionados con ellos, inconsistencia en información periférica, recuperación de nueva información poco relevante para los hechos). Por su parte, el sujeto mentiroso narra historias aprendidas con lo que las repetirá básicamente igual guiado por un esquema episódico. Es importante tener presente que esta segunda declaración siempre debe ser vista desde la óptica de que no contribuya significativamente a una victimación secundaria.

c) *Contraste de las declaraciones recabadas en el proceso judicial*. Asimismo, se analizan, de acuerdo con el procedimiento de estudio de la validez del SRA y del SVA, las otras declaraciones hechas a lo largo del proceso judicial (v. gr., sumario, diligencias). Ahora bien, el valor de éstas es relativo. Es preciso tener en mente que muchas de ellas son transcripciones de lo que un testigo ha dicho con lo cual no reflejan fehacientemente lo testificado. Además, el tipo de interrogatorio puede haber mediatizado la respuesta. Al respecto han de observarse, en línea con el SVA, los efectos sobre la validez de la declaración de las características de la entrevista (tipo de preguntas formuladas y adecuación de la entrevista) y motivación (motivos, contexto y presiones). Por ejemplo, en el caso de interrogatorios a menores hemos encontrado muchas expresiones y conceptos que preguntado el menor no sabe qué son (v. gr., en la declaración del menor aparece la expresión *echó semen*, si preguntado el menor qué es *semen*, y éste no lo sabe, es que no responde a su declaración tal expresión). A su vez, las declaraciones se suelen referir a expresiones (v. gr., me violó) no a narraciones de hechos (esto es, un evento narrativo de los hechos) con lo que no se puede contrastar su fiabilidad y validez. Por tanto, la falta de consistencia de las declaraciones obtenidas ante los peritos y otras re-

cogidas en el sumario tiene un valor muy relativo. En su caso, debe explicarse que esta falta de consistencia no es relevante para el análisis de la plausibilidad de la declaración. Por otra parte, es importante tomar con más precauciones de las que podrían esperarse a priori las confesiones por parte del acusado, y, muy especialmente, de las incriminaciones a cambio de beneficios para el informador. La fuente de sesgo viene de la mano de los interrogatorios. Así, las técnicas habituales para conseguir una confesión se basan en estrategias tales como las amenazas; la atribución de responsabilidad a causas externas como la provocación por parte de la víctima; minimización de la seriedad del crimen; o el desarrollo de una relación personal con el sospechoso (esto es, la típica estrategia de dos entrevistadores uno hostil, y otro amigable y protector). Finalmente, la estrategia basada en el dilema del prisionero para la obtención de la declaración puede llevar bien a estrategias de cooperación bien de competición que distorsionan la emisión del testimonio (v. gr., Kelley y Stahelski, 1970). Al respecto, una decisión de la Corte Suprema de los Estados Unidos (Miranda v. Arizona, 1966) declaró este tipo de interrogatorios coercitivos.

- d) *Análisis de contenido de las declaraciones referidas a los hechos.* El análisis de contenido de las declaraciones se dirige a dos dimensiones: la validez y la fiabilidad del testimonio. De acuerdo con el *Sistema de Evaluación Global*, el primer cometido de la evaluación pericial consiste en la estimación de validez de la declaración no como prueba judicial en sí, que corresponde a los Órganos Judiciales, sino como prueba para ser sometida a análisis de la fiabilidad de la misma. Al respecto, los ataques a la validez pueden ser dos. Primero, la declaración puede ser de tamaño insuficiente para ser objeto de un análisis de la realidad (Raskin y Steller, 1989). Segundo, la declaración puede resultar prueba no válida porque así lo informen los criterios de validez del SRA y SVA (p. e., carencia de consistencia interna; carencia de consistencia externa con otras pruebas robustas o incontrovertibles, tal como las periciales, recabadas en el proceso judicial; declaración inconsistente con la anterior; carencia de persistencia en las declaraciones, inconsistencia con las leyes científicas y de la naturaleza) y, en el caso de menores, indicios que limitan la validez (indicios

de sugestión, inadecuación del afecto, inadecuación del lenguaje y conocimientos). De resultar prueba no válida, se concluye que las declaraciones no constituyen prueba admisible o suficiente, y, de ser válida, se pasa al análisis de la fiabilidad (consistencia con los criterios de realidad) de las declaraciones. Como listado categorial de referencia tomamos las categorías del CBCA. Este procedimiento de análisis, creado en principio para el testimonio de menores víctimas de agresiones sexuales, es igualmente efectivo con adultos (Landry y Brigham, 1992; Zaparnuik, Yuille y Taylor, 1995; Spörer, 1997; Vrij, Edward, Roberts y Bull, 1999), en secuencias de medidas, y en otros casos diferentes a la agresión sexual (Porter y Yuille, 1996; Spörer, 1997; Arce, Fariña y Freire, 2002). En estos nuevos contextos obviamente no todas las categorías son productivas. Así, Landry y Brigham (1992) limitan el uso a 14 categorías con adultos porque tres de ellas sólo son aplicables a menores (incomprensión de detalles relatados con precisión; perdón al autor del delito; y detalles característicos de la ofensa) mientras que otras dos (elaboración inestructurada y asociaciones externas relacionadas) no eran productivas. Sin embargo, nosotros (Arce, Fariña y Freire, 2002) encontramos que la categoría *perdón al autor del delito* era productiva, en declaraciones de adultos, tanto en agresiones sexuales como en amenazas, esto es, la productividad está sujeta a un efecto de contexto. En suma, en principio deben considerarse todos los criterios en el análisis porque la productividad depende del tipo de caso, de las particularidades de la acción a examinar y del perfil sociodemográfico del entrevistado. A su vez, la combinación de los criterios del CBCA y RM es posible y efectiva, ya que pueden sumar sus efectos (Spörer, 1997; Vrij y otros, 1999). En concreto, la combinación de ambos sistemas de evaluación, resultando de añadir al CBCA los criterios *información perceptual* y *operaciones cognitivas* del RM (Vrij, 2000), mejora ligeramente la fiabilidad del sistema. Por tanto, procede a la inclusión de estos dos nuevos criterios a añadir a los del CBCA. Este procedimiento se puede aplicar en medidas repetidas (véanse las hipótesis a contrastar en el apartado *repetición de la obtención de la declaración*).

- e) *Análisis de la fiabilidad de las medidas.* Los sistemas originales de análisis de contenido de las declaraciones constituyen unas técnicas semi-objetivas por-

que sólo someten a contraste la fiabilidad y validez de los instrumentos, pero no contienen procedimientos de control de la medida concreta, esto es, de la medida pericial. Para subsanar esta laguna metodológica y acercarnos a un sistema objetivo, proponemos proceder con una metodología que posibilite verificar la fiabilidad de la medida a través del análisis de la consistencia inter- e intra-medidas, inter-evaluadores e inter-contextos (Wicker, 1975). La fiabilidad inter-contexto se afronta recurriendo a un evaluador entrenado que haya sido efectivo y consistente en otros contextos previos, o sea, en pericias anteriores. El recurso a dos evaluadores, habiendo sido al menos uno de ellos entrenado y fiable en evaluaciones anteriores, que ejecuten la tarea por separado posibilita obtener una aproximación a la consistencia inter-evaluadores. Como herramienta estadística de análisis de la consistencia inter-evaluadores proponemos el índice de concordancia [ $IC = \frac{\text{Acuerdos}}{(\text{acuerdos} + \text{desacuerdos})}$ ], que es más restrictivo que los valores kappa, tomando como punto de corte ,80 (Tversky, 1977). En otras palabras, sólo se consideran los resultados fiables si dos evaluadores, por separado, concuerdan en más de ,80 del total de evaluaciones en cada categoría de análisis. El contraste de la consistencia inter- e intra-medidas se aborda a través de la consistencia interna de las medidas (p. e., las escalas de validez del MMPI, de las declaraciones o del estudio de las estrategias de simulación en la entrevista clínica), de la consistencia entre distintas medidas (v. gr., concordancia entre MMPI y entrevista clínica, de las declaraciones en el tiempo) así como de la consistencia, esto es, complementación o no (verbigracia, una presenta indicios de veracidad y la otra de falsedad o están ausentes) de las evaluaciones obtenidas del denunciante y denunciado.

- f) *Medida de las secuelas clínicas del hecho traumático.* El acto delictivo produce una serie de daños en la víctima que básicamente son de orden físico, psíquico y económico. El daño psíquico conforma la denominada huella psíquica del delito y, como tal, puede aportarse como prueba de cargo. En relación con la valoración del daño psíquico y la subsecuente prueba judicial, los actos delictivos (p. e., lesiones, allanamientos de morada, malos tratos, abusos sexuales, secuestros) pueden producir una respuesta psicológica que se corresponde con el diagnóstico del trastorno de estrés posttraumático

(TEP) (Blanchard y Hickling; 2004, Echeburúa y Corral, 1998; Echeburúa, Corral, Zubizarreta y Sarasúa, 1995). Por ello, la medida del TEP es clave para la detección de la huella psicológica. Especial cuidado debe tenerse con las medidas indirectas del TEP (p. e., hipocondriasis, histeria, depresión, ansiedad, distimia, aislamiento social, inadaptación social), que pueden servir como potenciadores del diagnóstico del TEP pero no sustituirlo. Además, es preciso descartar otras causas al margen del acto delictivo. Por ejemplo, la unión de un proceso de separación con malos tratos, puede no permitir diferenciar cuál es el origen del trastorno, dado que ambas contingencias producen una huella psicológica semejante. En todo caso, el perito ha de tener en cuenta las siguientes máximas: no todos los actos delictivos producen un TEP en la víctima; y la no presencia del TEP no implica que la agresión no haya ocurrido. Al mismo tiempo, una vez identificada la huella psicológica del daño psíquico es preciso controlar que ésta no sea producto de una simulación por lo que se sigue el protocolo anteriormente descrito para la medida de la huella psíquica con control de la simulación.

- g) *Evaluación de la declaración de los actores implicados.* Si bien en un principio las técnicas de análisis de contenido de las declaraciones y la evaluación del daño psíquico fueron creadas para la valoración del testimonio del denunciante, el mismo procedimiento de análisis de contenido de las declaraciones también es susceptible de aplicación al denunciado, lo que permite llevar a cabo un estudio de las dos versiones. El procedimiento de justicia inquisitorial, como es el nuestro, permite esta doble confrontación, no así un sistema de adversarios. Con este procedimiento obtendremos una estimación de la validación convergente de los datos.
- h) *Análisis de la personalidad y capacidades de los actores implicados.* El estudio de la personalidad de las partes puede ser de suma trascendencia para explicar bien la denuncia, bien las claves explicativas de la agresión o cualquier enfermedad mental del denunciado con implicaciones jurídicas relevantes, o sea, se somete a estudio, dado el caso, la imputabilidad del encausado [en Arce, Fariña y Pampillón (2002) puede verse como se lleva a cabo el estudio de la imputabilidad]. Como quiera que la evaluación clínica no es suficiente en el contexto forense, se sigue el protocolo de evaluación clínica

con control de la simulación de Arce, Fariña y Pampillón (2002). Por su parte, las capacidades cognitivas de los actores se miden a través de la Escala Wechsler correspondiente y como fuente de contraste o para muestras con dificultades de lenguaje, mal escolarizadas o escolarizadas fuera del territorio español, se toma el test de inteligencia no verbal TO-NI-2 (Brown, Sherbenou y Johnsen, 1995). La fiabilidad de esta última medida se coteja a través de la correspondencia de las respuestas con el gradiente de dificultad de las preguntas y la consistencia inter-medidas. De la evaluación de las capacidades cognitivas nos servimos para informar de la capacidad de testimoniar y, en su caso, de sus efectos en la responsabilidad criminal.

- i) *Implicaciones para la presentación del informe.* El sistema de la credibilidad de las declaraciones en 5 categorías de respuesta, tal y como se recoge en el SVA, no se ajusta a los requerimientos de nuestro sistema de justicia. Así, el TS exige la seguridad plena, no la alta probabilidad (p. e., sentencia del TS de 29 de octubre de 1981, RA 3902). No obstante, toda medida, y muy especialmente la psicológica, está sujeta a error, por lo que debemos reconocerlo, pero absteniéndonos de establecer grados de certeza que, de acuerdo con las consideraciones del TS, sólo conllevan a una mayor confusión. De este modo, las categorías más ajustadas serían *probablemente cierto*, *probablemente no-cierto* y, en su caso, *indeterminado* (el perito psicólogo interesado podrá obtener de los autores un formato de pericia basado en el SEG). Téngase presente también que el sistema es más robusto en la identificación de la verdad que de la mentira. Asimismo, no es aconsejable que se haga una descripción de los hechos basada en frases, sino en acciones de conjunto, porque el procedimiento valida hechos y no partes aisladas. Por ello, en ningún caso se recomienda identificar al agresor porque el procedimiento no valida tal punto.

### CONSIDERACIONES FINALES

La fiabilidad de todo el procedimiento recae, en última instancia, en el entrevistador/evaluador. Es por ello que se hace preciso que la intervención se realice por profesionales con alta formación y experiencia, así como con una alta capacidad de objetividad (Alonso-Quecuty, 1993). Por eso, es imprescindible un entrenamiento exhaustivo. Éste debe incluir: a) entrenamiento en los mo-

dos de obtención de la información en todas sus modalidades (un procedimiento puede verse en Fisher y otros, 1987); b) entrenamiento en análisis de las declaraciones [un programa estructurado puede verse en Köhnken (1999)]; c) entrenamiento en evaluación de la personalidad y la huella psíquica, no con fines clínicos, sino forenses (véase Arce, Fariña y Freire, 2002; Arce, Fariña y Pampillón, 2002; Echeburúa, Corral y Amor, 2002; Rogers, 1997); d) entrenamiento en la detección de la simulación (véase en Arce, Fariña y Pampillón, 2002 los pasos a seguir); y e) ejecución de las primeras evaluaciones forenses en compañía de un perito con experiencia. Finalmente, nuestra experiencia nos dice que el material de entrenamiento en análisis de contenido y evaluación clínica ha de ser real y no simulaciones pues la tarea ejecutada en estos contextos es distinta y la eficacia del procedimiento también (Vrij, 2000). La Unidad de Psicología Forense de la Universidad de Santiago de Compostela organiza periódicamente cursos de entrenamiento en estas técnicas.

**Agradecimientos:** Esta investigación fue financiada por la Secretaría Xeral de Investigación e Desenvolvemento, Xunta de Galicia, en el proyecto de Excelencia Investigadora Código PGDIT03CS037401PR.

### REFERENCIAS

- Alonso-Quecuty, M. L. (1993). Interrogando a testigos, víctimas y sospechosos: La obtención de información exacta. En M. Diges, y M. L. Alonso-Quecuty (eds.), *Psicología forense experimental* (pp. 85-98). Valencia: Promolibro.
- American Psychiatric Association, (2002). *Manual diagnóstico y estadístico de los trastornos mentales* (IV Edición-Texto Revisado). Barcelona: Masson.
- Arce, R. y Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de delitos: La entrevista clínico-forense*. Manuscrito Inédito, Universidad de Santiago de Compostela.
- Arce, R. y Fariña, F. (2002). *El Sistema de Evaluación Global para la evaluación de la credibilidad y el engaño en el testimonio judicial*. Manuscrito Inédito, Universidad de Santiago de Compostela.
- Arce, R. y Fariña, F. (2005). El Sistema de Evaluación Global (SEG) de la credibilidad del testimonio: Hacia una propuesta integradora. En R. Arce, F. Fariña, y M. Novo (eds.). *Psicología jurídica* (pp. 101-118).



- Santiago de Compostela: Consellería de Xustiza, Interior e Administración Local.
- Arce, R. y Fariña, F. (en prensa). Evaluación de la credibilidad y de la huella psíquica en el contexto penal. En Consejo General de Poder Judicial (ed.), *Psicología del testimonio y valoración de la prueba pericial*. San Sebastián: Centro de Documentación Judicial.
- Arce, R., Carballal, A., Fariña, F. y Seijo D. (2004). Can mock battered women malingering psychological evidence in a recognition task? En A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch y J. Wójcikiewicz (eds.), *Forensic psychology and law. Facing the challenges of a changing world* (pp. 327-336) Kraków: Institute of Forensic Research Publishers.
- Arce, R., Fariña, F. y Freire, M. J. (2002). Contrastando la generalización de los métodos empíricos de detección del engaño. *Psicologia: Teoria, Investigação e Prática*, 7(2), 71-86.
- Arce, R., Fariña, F. y Real, S. (2000). The assessment of information integration theory and confirmatory bias hypothesis on judicial proceedings. A case of rape and murder. En A. Czerederecka, T. Jaskiewicz-Obydzinska y J. Wójcikiewicz (eds.), *Forensic psychology and law* (pp. 296-303) Kraków: Institute of Forensic Research Publishers.
- Arce, R., Fariña, F. y Seijo, D. (2003). Laypeople's criteria for the discrimination of reliable from non-reliable eyewitnesses. En M. Vanderhallen, G. Vervaeke, P. J. Van Koppen y J. Goethals (eds.), *Much ado about crime* (pp. 105-116). Bruselas: Uitgeverij Politeia NV.
- Arce, R., Fariña, F., Carballal, A. y Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación. *Psicothema*, 18.
- Arce, R., Fariña, F., Novo, M. y Seijo, D. (2001). Judges' decision-making from within. En R. Roesch, R. R. Corrado y R. J. Dempster (eds.), *Psychology in the courts: international advances in knowledge* (pp. 195-206). Nueva York: Routledge.
- Arce, R., Novo, M. y Alfaro, E. (2000). La obtención de la declaración en menores y discapacitados. En A. Ovejero, M. V. Moral y P. Vivas (eds.), *Aplicaciones en psicología social* (pp. 147-151). Madrid: Biblioteca Nueva.
- Arce, R., Pampillón, M. del C. y Fariña, F. (2002). Desarrollo y evaluación de un procedimiento empírico para la detección de la simulación de enajenación mental en el contexto legal. *Anuario de Psicología*, 33(3), 385-408.
- Bagby, R. M., Buis, T. y Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: replication and extension. *Psychological Assessment*, 7, 84-92.
- Bagby, R. M., Rogers, R. Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., Schuller, D. R. y Seeman, M. V. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, 68(3), 650-664.
- Bauer, P. J. y Mandler, J. M. (1990). Remembering what happened next: Very young children's recall of event sequences. En R. Fivush, y J. A. Hudson (eds.), *Knowing and remembering in young children* (pp. 9-29). Nueva York: Cambridge University Press.
- Blanchard, E. B. y Hickling, H. J. (2004). What are the psychosocial effects of MVAs on survivors? En E. B. Blanchard, E. J. Hickling (eds.), *After the crash: Psychological assessment and treatment of survivors of motor vehicle accidents* (2ª Ed., pp. 57-97). Washington, DC: American Psychological Association.
- Bower, G. (1967). A multicomponent theory of memory trace. En K. W. Spence y J. T. Spence (eds.), *The psychology of learning and motivation* (Vol. 1) (pp. 229-325). Nueva York: Academic Press.
- Bower, G. H. y Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247, 44-48.
- Brown, L., Sherbenou, R. y Johnsen, S. (1995). *Toni-2*. Madrid: TEA Ediciones.
- Bull, R. (1995). Interviewing people with communicative disabilities. En R. Bull y D. Carson (eds.), *Handbook of psychology in legal contexts* (pp. 247-260). Chichester: John Wiley and Sons.
- Bull, R. (1997). Entrevistas a niños testigos. En F. Fariña, y R. Arce (eds.), *Psicología e investigación judicial* (pp. 19-38). Madrid: Fundación Universidad Empresa.
- Butcher J. N. y Miller, K. B. (1999). Personality assessment in personal injury litigation. En A. K. Hess e I. B. Weiner (eds.), *The handbook of forensic psychology* (2ª Ed., pp. 104-126). Nueva York: John Wiley and Sons.
- Campos, L. y Alonso-Quecuty, M. L. (1999). The cognitive interview: Much more than simply "try again". *Psychology, Crime and Law*, 5(1-2), 47-59.
- Cunniën, A. J. (1997). Psychiatric and medical syndromes associated with deception. En R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 23-46). Nueva York: Guilford Press.
- Dammeyer, M. D. (1998). The assessment of child sexual abuse allegations: Using research to guide clinical de-

- cision making. *Behavioral Sciences and the Law*, 16, 21-34.
- Davies, G., Tarrant, A. y Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology*, 80, 415-429.
- Derogatis, L. R. (2002). *SCL-90-R. Manual*. Madrid: TEA Ediciones.
- Duckworth, J. y Anderson, W. P. (1995). *MMPI and MMPI-2: Interpretation manual for counsellors and clinicians*. Bristol: Accelerated Development.
- Echeburúa, E. y Corral, P. (1998). *Manual de violencia familiar*. Madrid: Siglo XXI.
- Echeburúa, E. y Corral, P. de y Amor, P. J. (2002). Evaluación del daño psicológico en víctimas de crímenes violentos. *Psicothema*, 14, 139-146.
- Echeburúa, E., Corral, P. de, Zubizarreta, I. y Sarasúa, B. (1995). *Trastorno de estrés postraumático crónico en víctimas de agresiones sexuales*. A Coruña: Fundación Paideia.
- Fisher, R. P. y Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interview*. Springfield: Charles C. Thomas.
- Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M. y Warhaftig, M. L. (1987). Enhancing eyewitness memory: Refining the cognitive interview. *Journal of Police Science and Administration*, 15, 291-297.
- Freed, D. M., Corkin, S., Growdon, J. H. y Nissen, M. J. (1989). Selective attention in Alzheimer's disease: Characterizing cognitive subgroups of patients. *Neuropsychologia*, 27, 325-339.
- Gil, R. (1999). *Neuropsicología*. Barcelona: Masson.
- Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M. y Kuhn, J. (1997). Children's reactions to and memory for a stressful event: Influences of age, anatomical dolls, knowledge, and parental attachment. *Applied Developmental Science*, 1, 54-75.
- Graham, J. R. (1992). *Interpretation of MMPI-2. Validity and clinical scales*. Brujas, workshop.
- Hart, R. P. y O'Shanick, G. J. (1993). Forgetting rates for verbal, pictorial, and figural stimuli. *Journal of Clinical and Experimental Neuropsychology*, 15, 245-265.
- Hathaway, S. R. y Mckinley, J. C. (1999). *MMPI-2. Inventario multifásico de personalidad de Minnesota-2. Manual*. Madrid: TEA Ediciones.
- Home Office and The Department of Health (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. Londres: HMSO.
- Johnson, M. K. y Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Kelley, H. H. y Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 16, 66-91.
- Köhnken, G. (1999). *Assessing credibility*. Pre-conference of the EAPL Programme of Applied Courses, Dublin.
- Köhnken, G., Milne, R., Memon, A. y Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, 5, 3-27.
- Landry, K. L. y Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, 16, 663-676.
- Lewis, M. y Saarni, C. (1993). *Lying and deception in everyday life*. Nueva York: Guilford Press.
- Loftus, E. F., Korf, N. L. y Schooler, J. W. (1988). Misguided memories: Sincere distortions of reality. En J. C. Yuille (ed.), *Credibility assessment* (pp. 155-172). Dordrecht: Kluwer.
- Mantwill, M., Köhnken, G. y Ascherman, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology*, 80, 68-78.
- Memon, A. y Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation and practical implications. *Journal of Community and Applied Social Psychology*, 1, 291-307.
- Memon, A., Cronin, O., Eaves, R. y Bull, R. (1993). The cognitive interview and the child witness. En N. K. Clark y G. M. Stephenson (eds.), *Issues in criminology and legal psychology: Vol. 20. Children, evidence and procedure*. Leicester: British Psychological Society.
- Memon, A., Cronin, O., Eaves, R. y Bull, R. (1996). An empirical test of mnemonic components of cognitive interview. En G. M. Davies, S. Lloyd-Bostock, M. McMurrin, y C. Wilson (eds.), *Psychology and law: Advances in research* (pp. 135-145). Berlín: Walter de Gruyter.
- Memon, A., Wark, L., Bull, R. y Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, 88, 179-198.
- Memon, A. y Bull, R. (1999). *Handbook of the psychology of interviewing*. Chichester: John Wiley and Sons.
- Mira, J. J. (1989). *Estudios de psicología en ambientes*

- procesales: *Un análisis de metamemoria*. Tesis Doctoral, Universidad Autónoma de Madrid.
- Miranda v. Arizona, 384 U.S. 436 (1966).
- Ostrom, T. M., Werner, C. y Saks, M. J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 36, 436-450.
- Peña-Casanova, J. (1990). *Programa integrado de exploración neuropsicológica*. Test Barcelona. Barcelona: Masson.
- Piñeiro, A. (2005). Criterios empíricos de credibilidad y profesionales de la justicia. En R. Arce, F. Fariña y M. Novo (eds.), *Psicología jurídica* (pp. 213-221). Santiago de Compostela: Consellería de Xustiza.
- Porter, S. y Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues in the interrogation context. *Law and Human Behavior*, 20(4), 443-458.
- Raskin, D. C. y Steller, M. (1989). Assessing credibility of allegations of child sexual abuse: Polygraph examinations and statement analysis. En H. Wegener, F. Lösel y J. Haisch (eds.), *Criminal behaviour and the justice system: Psychological perspectives* (pp. 290-302). Heidelberg: Springer-Verlag.
- Rogers, R. (1992). *Structured interview of reported symptoms*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R. (1997). Researching dissimulation. En R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 398-426). Nueva York: Guilford Press.
- Rogers, R. y Mitchell, C. N. (1991). *Mental health experts and the criminal courts: A handbook for lawyers and clinicians*. Scarborough, ON: Thomson Professional Publishing Canada.
- Roig-Fusté, J. M. (1993). *MMPI y MMPI-2 en la exploración de la personalidad*. Barcelona: Agil Offset.
- Sallmann, P. y Willis, J. (1984). *Criminal justice in Australia*. Melbourne: Oxford University Press.
- Schum, D. A. (1977). The behavioral richness of cascaded inference models: Examples of jurisprudence. En N. J. Castellan, D. B. Pisoni y G. R. Potts (eds.), *Cognitive theory* (vol. 2). Hillsdale, N.J.: LEA.
- Spörer, S. L. (1997). The less traveled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373-397.
- Steller, M. (1989). Recent developments in statement analysis. En J. C. Yuille (ed.), *Credibility assessment*. (pp. 135-154). Dordrecht: Kluwer Academic Publishers.
- Steller, M. y Köhnken, G. (1994). Análisis de declaraciones basados en criterios. En D. C. Raskin (ed.), *Métodos psicológicos en la investigación y pruebas criminales* (pp. 217-245). Bilbao: Desclée de Brouwer (Orig. 1989).
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.
- Tulving, E. y Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 353-370.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Zeugenaussagen. En U. Undeutsch (ed.), *Handbuch der psychologie, Vol. II: Forensische psychologie* (pp. 26-181). Göttingen: Verlag für Psychologie.
- Undeutsch, U. (1988). The development of statement reality analysis. En J. Yuille (ed.), *Credibility assessment* (pp. 101-119). Dordrecht: Kluwer.
- Vrij, A. (2000). *Detecting lies and deceit*. Chichester: Wiley.
- Vrij, A. y Winkel, F. W. (1996). Detection of false statements in first and third graders: The development of a nonverbal detection instrument. En G. Davies, S. Lloyd-Bostock, M. McMurray y C. Wilson (eds.), *Psychology, Law, and Criminal Justice* (pp. 221-230). Berlin: Walter de Gruyter.
- Vrij, A., Edward, K., Roberts, K. P. y Bull, R. (1999, julio). *Detecting deceit via criteria based content analysis, reality monitoring and analysis of non verbal behavior*. The First Joint Meeting of the American Psychology-Law Society and the European Association of Psychology and Law, Dublin.
- Wechsler, D. (1976). *WAIS: Escala de inteligencia para adultos (2ª ed.)*. Manual. Madrid: TEA Ediciones.
- Weick, K. E. (1985). Systematic observational methods. En G. Lindzey y E. Aronson (eds.), *The handbook of social psychology* (vol. 1, pp. 567-634). Hillsdale, N. J.: LEA.
- Wicker, A. W. (1975). A application of a multiple-trait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.
- Winograd, E., Smith, A. D. y Simon, E. W. (1982). Aging and the picture superiority effect in recall. *Journal of Gerontology*, 37, 70-75.
- Zaparniuk, J. Yuille, J. C. y Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, 18, 343-352.

## ¿SE PILLA ANTES A UN MENTIROSO QUE A UN ÇOJO? SABIDURÍA POPULAR FRENTE A CONOCIMIENTO CIENTÍFICO SOBRE LA DETECCIÓN NO-VERBAL DEL ENGAÑO

### IS THE LIAR CAUGHT SOONER THAN THE CRIPPLE? POPULAR WISDOM VERSUS SCIENTIFIC KNOWLEDGE ON THE NON-VERBAL DETECTION OF DECEPTION

Jaume Masip

Universidad de Salamanca

*Existe una serie de creencias populares sobre los indicadores conductuales del engaño que no se ven corroboradas por la evidencia empírica. Determinados libros "de autoayuda" contribuyen a la difusión de las mismas. En este trabajo se revisan varias décadas de investigación en psicología y comunicación sobre la detección no-verbal del engaño. Al contrario de lo que propugnan los libros "de autoayuda" y de lo que sostiene la sabiduría popular, detectar la mentira a partir del comportamiento no-verbal es extremadamente difícil, apenas sí existen claves conductuales que permitan discriminar entre verdades y mentiras, su significado y poder de discriminación varían en función de diversas variables contextuales, y la eficacia de los programas de entrenamiento es muy limitada. Frente a las cuestionables afirmaciones de determinados libros populares y dadas las graves consecuencias que en ciertos ámbitos pueden tener los juicios de credibilidad erróneos, es necesario desmontar los falsos mitos existentes sobre la detección no-verbal de la mentira, sustituyéndolos por información más válida y científicamente contrastada.*

*There is a series of popular beliefs about the behavioral indicators of deception that are not supported by empirical research. A number of "self-help" books are contributing to the spreading around of these beliefs. In this article, several decades of psychological and communication research on the nonverbal detection of deception are reviewed. Contrary to the claims of "self-help" books and to the tenets of popular wisdom, detecting deception from behavioral cues is extremely difficult, there are almost no behavioral cues to differentiate between truths and lies, their meaning and usefulness are under the influence of a number of contextual variables, and training programs have yielded only very limited improvements in accuracy. In view of the misleading contents of certain popular books and the serious consequences of wrong credibility judgments in a number of contexts, it is necessary to dispel the existing false myths about the nonverbal detection of deceit, providing instead valid and scientifically tested information.*

La psicología social ha abordado en diversas ocasiones la relación entre el conocimiento psicológico y el sentido común (por ej., Garrido, Herrero y Masip, 2004; Teigen, 1986; véanse las consideraciones de Kelley, 1992). Como han señalado algunos autores (por ej., Myers, 1999/2000), se critica la psicología social porque estudia cosas que todo el mundo sabe, es decir, que son "de sentido común" (Kelley, 1992; Schlesinger, 1949). Esta crítica se formula, casi siempre, una vez se ofrece al crítico la respuesta correcta ("¡esto ya lo sabía yo!"); pero por lo general le cuesta adivinar de antemano tan "obvia" respuesta (Kelley, 1992; Lazarsfeld, 1949).

Hay dos ámbitos por los que siento un interés profesional especial desde hace algunos años en los que la distancia entre el sentido común y los hallazgos científicos

es notable. Tales ámbitos son el de la comunicación no-verbal y el de la detección del engaño. Probablemente por su atractivo intrínseco, ambos estimulan la imaginación popular, dando lugar a "teorías" y visiones de lo más extravagantes que, por lo general, no tardan en ganar la más amplia aceptación popular. A esta difusión suelen contribuir un sinnúmero de libros oportunistas, mal llamados "de autoayuda", firmados, en muchos casos, por profesionales de formación dudosa que se aprovechan de la ingenuidad del lector, movidos probablemente por intereses económicos, obrando con ello de modo éticamente reprochable. No pretendo cuestionar todos los libros de autoayuda. Ciertamente, algunos de ellos son obra de reputados investigadores y están escritos con absoluto rigor científico. Pero buena parte de las publicaciones de este tipo se basan en las creencias ingenuas y sin contrastar de sus autores, y no en el estado de conocimiento científico sobre el tema. Ello hace escaso bien a la difusión popular del conocimiento, y no contribuye a la auto-ayuda que pueda ofre-

*Correspondencia:* Jaume Masip. Departamento de Psicología Social y Antropología. Universidad de Salamanca. Facultad de Psicología. Avda. de la Merced, 109-131. 37005 Salamanca. España. E-mail: jmasip@usal.es



cerse a sí mismo el interesado lector (razón por la que se entrecomilla en este trabajo el calificativo “de autoayuda” referido a tales libros). Resulta curioso que, precisamente, sea el desconocimiento que el lector tiene de la disciplina el factor que, por una parte, le impulsa a acudir a esas publicaciones, mientras que por otra parte le impide apreciar el escaso valor científico de las mismas, haciéndole vulnerable a la desinformación que éstas destilan.

Específicamente en relación con el ámbito de la comunicación no-verbal (o “lenguaje del cuerpo”, como se suele llamar en tales libros) se da la circunstancia de que, tal como ocurre en otros campos de la psicología, todo el mundo “sabe” sobre el tema, y todo el mundo opina al respecto, osando negar, relativizar o matizar la palabra del verdadero experto. Es como si los estereotipos populares tuvieran más valor que el conocimiento científico obtenido mediante los rigurosos procedimientos aceptados en el ámbito de las ciencias. A menudo uno se sorprende al encontrar publicidad en que se anuncian cursillos de “comunicación exitosa”, o con sensacionalistas títulos como “el lenguaje del cuerpo para las ventas” o similares, ofrecidos por consultorías u otros organismos ajenos a nuestro ámbito de especialización, e impartidos por personas cuyos escasos conocimientos en psicología o comunicación interpersonal no les capacitan en absoluto para impartir esos contenidos. Sería absurdo que un psicólogo pretendiera ofrecer un cursillo sobre derecho, economía o ingeniería. Sin embargo, algunos abogados, economistas, ingenieros y profesionales de otros “gremios” alejados del nuestro no dudan en considerarse capacitados para adentrarse sin vacilar en el campo de la psicología, la comunicación y otras ciencias sociales afines para impartir “conocimientos” especializados. En mi opinión, esto merece el calificativo de intrusismo profesional.

El resultado de tal estado de cosas no puede sino ser la difusión de falsas creencias sobre el significado del comportamiento, así como ofrecer la imagen distorsionada de que la conducta no-verbal es un “juego de niños”, con gestos de significado inequívoco y carente de todo relativismo. Así, por ejemplo, muchos creen que cruzar las piernas o los brazos significa indudablemente que la persona no está “abierta” psicológicamente al otro, que proyectar la mandíbula hacia delante es un signo de dominancia, etc. Uno no puede sino sonreírse ante la ingenuidad de tales creencias, que reflejan más una serie de

teorías implícitas a menudo equívocas que el verdadero conocimiento científico de nuestra disciplina.

Un buen ejemplo de la difusión de este tipo de creencias lo constituye el popular libro *El lenguaje del cuerpo*, de Allan Pease (1981/1988). El autor, un vendedor a comisión, empezó a interesarse por el “lenguaje del cuerpo” al asistir a un seminario ofrecido en 1971 por el antropólogo Ray Birdwhistell. Resulta desafortunado que, en su obra, Pease no haga honor a la indiscutible reputación científica de Birdwhistell, pese a la engañosa afirmación que incluye en el prólogo de que “en este libro he resumido algunos de los estudios realizados por los mejores especialistas en el comportamiento humano” (Pease, 1988, p. 9).

Si bastante dañina es ya la difusión de falsas creencias “disfrazadas” de conocimiento científico por legos en la materia, el tema adquiere tintes escandalosos cuando quienes las difunden son supuestos profesionales. Paolo Abozzi, que se erige en director de un llamado *Centro di Comunicazione Integrale* en Roma y que afirma tener formación en comunicación e hipnosis (véase <http://digilander.libero.it/magopaolo/PAOLO%20ABOZZI.html>), es autor de, entre otras obras, el libro *La interpretación de los gestos* (Abozzi, 1996/1997). La naturaleza del mismo es idéntica a la del volumen de Pease, y lo cierto es que el *Centro di Comunicazione Integrale* no es, como en principio podría pensarse, un centro de investigación, sino un organismo que se dedica a ofrecer cursillos y vídeos sobre hipnosis, grafología, programación neurolingüística y temáticas similares (<http://digilander.libero.it/magopaolo/index2.html>). El peligro que supone la difusión de falsos conocimientos por supuestos profesionales se fundamenta en el conocido impacto de la credibilidad de la fuente sobre la persuasión (Kruglanski *et al.*, 2005). El cliente ingenuo va a considerar ciertas esas informaciones por provenir de un “experto” en el tema, por lo que creará a pies juntillas todas sus aseveraciones y seguirá las recomendaciones ofrecidas. Ello puede llevar a decisiones erróneas de graves consecuencias en contextos como el interpersonal, el laboral o el jurídico.

El segundo ámbito al que hacía referencia anteriormente es el de la detección de la mentira. Siendo tan “intrigante” como el del comportamiento no-verbal, se ve amenazado por idénticos peligros. Estos se han concretado, por ejemplo, en diversas técnicas o procedimientos elaborados por veteranos policías o militares cuya experiencia profesional



en contextos en que la mentira es frecuente les dota de cierta credibilidad popular<sup>1</sup>. Pero el que un profesional tenga *experiencia* no implica necesariamente que deba ser un *experto* (en este sentido, y aplicado específicamente al ámbito de la detección no-verbal del engaño, véanse los trabajos de DePaulo y Pfeiffer, 1986; Garrido, Masip y Herrero, 2004; Meissner y Kassin, 2002; Strömwall, Granhag y Hartwig, 2004). En consecuencia, sus recomendaciones pueden ser equivocadas. El auge de aparatos tales como los evaluadores del estrés vocal (Masip, Garrido y Herrero, 2004) o de procedimientos como la Técnica SCAN (Masip, Garrido y Herrero, 2002a) son claros ejemplos de lo dicho. Desarrollados por veteranos profesionales de los cuerpos de seguridad, tales artilugios y procedimientos gozan de gran popularidad en ámbitos aplicados, debido en parte a la profesión de sus creadores y en parte a los poderosos mecanismos de marketing desarrollados a su alrededor. Sin embargo, su utilidad real para detectar mentiras ha sido seriamente cuestionada por la investigación empírica. El riesgo es, una vez más, la gravedad de las consecuencias derivadas del uso de la información errónea proporcionada. Si el mito de que los evaluadores del estrés vocal o la Técnica SCAN son instrumentos o procedimientos válidos y fiables está bien enraizado en la sociedad, probablemente la judicatura admita como prueba en los juicios la evidencia obtenida con ellos. Pero si en realidad no discriminan adecuadamente entre personas veraces y mentirosas se puede estar condenando de forma injusta a sospechosos inocentes, al tiempo que los verdaderos culpables quedan en libertad (en este sentido, véase el informe del National Research Council, 2003, referido al empleo de polígrafo).

Si el problema ya resulta de consideración al tomar en cuenta separadamente el comportamiento no-verbal y la detección de la mentira, no debe sorprender que la situación sea poco alentadora cuando se trata de detectar la mentira a partir del comportamiento no-verbal. Hace unos años encontré anunciado en un catálogo un libro escrito por alguien llamado David Lieberman (1998) que llevaba por título *Never be lied to again* (algo así como "Que no te mientan nunca más"). Lo solicité, si bien con abierto escepticismo dada la naturaleza sensacionalista del título y mi absoluto desconocimiento del autor (era evidente que no se trataba de ningún investi-

gador relevante en esta área). El libro, subtítulo "cómo saber la verdad en 5 minutos o menos en cualquier conversación o situación", no contiene información alguna con valor científico o práctico, sino una colección de absurdos consejos de naturaleza completamente engañosa. Lo más indignante del caso son las letras "Ph.D." que figuran en la cubierta y el lomo del volumen junto al nombre del autor, y que designan que éste es Doctor. Asimismo, en la solapa se ensalzan las supuestas virtudes profesionales de dicho autor. No soy contrario a que cada cual se exprese libremente, escribiendo las más fantasiosas excentricidades; pero otra cosa muy distinta es intentar hacer pasar por información científica y contrastada (como evidencia la inclusión de las letras "Ph.D." y los datos de la solapa del libro) una serie de contenidos de ningún valor. Se trata, simple y llanamente, de un fraude, y deberían emprenderse acciones legales contra fraudes de esta naturaleza. Sólo queda esperar que ningún profesional (policía, juez, abogado, etc.) de cuyas decisiones sobre la sinceridad de otra persona dependa el destino de ésta lea el libro o se lo tome con seriedad.

Un ejemplo dramático de las consecuencias prácticas que puede tener la difusión de datos o procedimientos científicos lo constituye el controvertido entrenamiento de Inbau, Reid, Buckley y Jane (2001). Impartido por la empresa *John E. Reid & Associates*, dicho entrenamiento se dirige a miembros de los cuerpos de seguridad que deban interrogar a sospechosos. La compañía se jacta de haber entrenado a más de 300.000 profesionales desde el primer seminario sobre interrogatorios y entrevistas celebrado en 1974 (véase <http://www.reid.com>). Parte del entrenamiento de Inbau *et al.* (2001) se centra sobre las claves conductuales del engaño. Sin embargo, las claves que se enseñan se apartan de aquellas pocas que la investigación empírica ha mostrado que pueden ser útiles (véase el interesante contraste ofrecido por Blair y Kooi, 2004). Asimismo, atender a tales indicadores reduce la precisión de los policías al juzgar la credibilidad de declaraciones verdaderas (Mann, Vrij y Bull, 2004). Además, Kassin y Fong (1999) han mostrado empíricamente que el entrenamiento en los indicadores de Inbau *et al.* produce una *reducción* en la precisión global alcanzada, acompañada de un sesgo a decir que

<sup>1</sup> Por ejemplo, Garrido, Masip y Herrero (2004) hallaron que se considera que los policías son más capaces de diferenciar entre verdades y mentiras que la población en general.

los sujetos mienten y de un incremento de la confianza en los juicios.

Si se tiene en cuenta que, en muchos países, antes de someter al sospechoso a un severo interrogatorio la policía sostiene una entrevista más distendida con él para establecer su inocencia o culpabilidad sobre la base de los indicadores conductuales del engaño, el peligro de la desinformación que proporciona *John E. Reid & Associates* es obvio. Pero este peligro se magnifica si se tiene en cuenta el tipo de interrogatorio que *John E. Reid & Associates* propone, pues se trata de una aproximación altamente agresiva y coercitiva que puede llevar a muchos inocentes a confesar el delito que se investiga (por ej., Kassin, 2005; Kassin y Gudjonsson, 2004). En definitiva, pues, la policía: (a) entrevista a un sospechoso; (b) observa determinados indicadores conductuales de escaso valor diagnóstico pero que cree asociados con el engaño, y en consecuencia resuelve que el sospechoso miente; y entonces (c) desde este convencimiento somete al sospechoso a un duro proceso de interrogatorio cuya naturaleza puede hacer que incluso muchas personas inocentes confiesen (Kassin, 2004, 2005; Kassin y Gudjonsson, 2004). Este proceso puede explicar buena parte del elevado número de casos, registrados en países como los Estados Unidos (donde la técnica de Inbau y Reid goza de cierta popularidad entre los miembros de las fuerzas de seguridad), de personas que han sido encarceladas sobre la base de una confesión que más tarde se ha demostrado fehacientemente que era falsa (Drizin y Leo, 2004).

El objetivo del presente trabajo consiste en “desmantelar” una serie de creencias populares erróneas, en muchas ocasiones difundidas a través de cursillos o libros escritos por personas poco cualificadas, referentes a un tema claramente “psicológico” como es la detección del engaño a partir del comportamiento no-verbal. La información que se proporciona en las siguientes páginas está basada en la más rigurosa investigación científica en psicología y comunicación interpersonal. Dicha información será de indudable interés para el profesional de la psicología debido a tres razones: ser parte de su disciplina, la utilidad que puede tener en muchos ámbitos aplicados de la misma, y por el papel asesor que el psicólogo desempeña al ser interpelado por otros profesionales, a cuyas consultas debe responder según la ciencia psicológica, cuestionando las creencias engañosas que pueda tener el que inquiere.

### PRECISIÓN: ¿SE PILLA ANTES A UN MENTIROSO QUE A UN COJO?

Una creencia popular muy extendida es la que se refleja en el dicho “se pilla antes a un mentiroso que a un cojo”. En otras palabras: es fácil pillar al mentiroso. ¿Es correcta esta creencia?

El examen de la precisión (nivel de aciertos) al hacer juicios de veracidad ha sido uno de los temas más estudiados en el área del engaño. El procedimiento experimental empleado suele consistir en presentar a una muestra de sujetos observadores o receptores una serie de declaraciones efectuadas por un conjunto de sujetos emisores (los potenciales mentirosos). Dichas declaraciones se presentan en formato audiovisual o auditivo, empleando grabaciones electromagnéticas o una representación “en vivo” (véase el Capítulo 3 de Miller y Stiff, 1993, para una descripción de los paradigmas experimentales empleados). En algunas ocasiones se permite que emisor y receptor interactúen libremente (Buller y Burgoon, 1996). Los receptores deben indicar, habitualmente en un formulario, si cada una de las declaraciones presentadas es verdadera o falsa. En ocasiones también se solicita de ellos que expresen el grado de confianza en cada juicio y los indicadores a los que han atendido para formular dicho juicio.

Normalmente la mitad de las declaraciones presentadas son verdaderas y la otra mitad son falsas. De modo que, sólo por azar, los observadores pueden acertar la mitad de los juicios, es decir, pueden tener una precisión del 50%. ¿Cuál es la precisión alcanzada realmente en los estudios empíricos? En 1980, Kraut publicó una revisión de los estudios realizados hasta el momento. En ella se indicaba que la precisión media era del 57%. Veinte años después, Vrij (2000) promedió la precisión obtenida en 39 estudios relevantes. Ésta fue casi idéntica a la hallada por Kraut: 56.6%. Aproximadamente un tercio ( $n = 12$ ) de los experimentos revisados por Vrij arrojaban una precisión que se situaba en el estrecho rango limitado por el 54% y el 56%. En ningún experimento la precisión estaba por debajo del 30% ni por encima del 64% (Vrij, 2000).

Más recientemente se han realizado revisiones mucho más exhaustivas y actualizadas, basadas en un muestreo más metódico de estudios. Aamodt y Mitchell (en prensa) han llevado a cabo un meta-análisis sobre el efecto de diversas variables individuales sobre la precisión al efectuar juicios de credibilidad. Examinando un total de

193 muestras distintas de receptores, con una cantidad total de 14.379 observadores, han hallado una precisión media del 54.5%. En otro trabajo más amplio (incluye un total de 349 muestras de receptores, con 22.282 sujetos que evaluaron la credibilidad de los mensajes de 3864 emisores), Bond y DePaulo (en prensa) hallaron una precisión media del 53.4%. Si bien ésta es significativamente superior al 50% esperado por azar, en términos absolutos es una precisión extremadamente pobre. Significa que de cada 100 mensajes hay 47 que se juzgan erróneamente. Es decir, tenemos casi la misma probabilidad de acertar nuestros juicios que de fallarlos. La precisión de los detectores humanos al hacer juicios de credibilidad sobre la base de la observación del comportamiento es, pese a lo que dice la sabiduría popular, extremadamente limitada. De hecho, de las diversas aproximaciones a la detección del engaño, la no-verbal es la que arroja unos niveles de precisión más bajos<sup>2</sup>.

Esta limitación se extiende asimismo a aquellos profesionales para los cuales detectar mentiras es importante y que tienen experiencia en tareas de evaluación de la credibilidad. Así, frente a la precisión del 54.2% obtenida por estudiantes universitarios legos, Aamodt y Mitchell (en prensa) informan de niveles del 50.8% para las muestras de detectives, del 54.5% para policías federales norteamericanos, del 55.3% para policías y para agentes de aduanas, del 59.0% para jueces y del 61.6% para

las cuatro muestras de psicólogos incluidas en su meta-análisis. Bond y DePaulo (en prensa) utilizan contrastes estadísticos para comparar la precisión de “expertos” (personal de los cuerpos de seguridad, jueces, psiquiatras, auditores...) y “no-expertos”. Ni en las comparaciones intraestudio (al considerar conjuntamente todos los experimentos en que se había hecho esta comparación) ni en las comparaciones interestudio (comparación del nivel de precisión en experimentos en que los observadores habían sido “expertos” con experimentos en que éstos habían sido “no-expertos”) las diferencias resultan significativas. En las comparaciones interestudio los niveles de precisión hallados han sido 52.9% para los “expertos” y 56.9% para los “no-expertos”. En definitiva, los profesionales familiarizados con el engaño *no* son mejores detectores que los observadores legos.

La precisión no sólo es baja, sino que además es *uniformemente* baja. Hay evidencia de que existe un conjunto de factores situacionales y personales que influyen de forma estadísticamente significativa sobre los juicios y los niveles de precisión (Masip, Garrido y Herrero, 2002b). Así, Bond y DePaulo (en prensa) hallaron que determinadas variables (canal de comunicación, motivación del emisor, preparación, exposición previa a la conducta del emisor e interacción vs. no-interacción emisor-receptor) tenían un impacto significativo sobre el nivel de aciertos<sup>3</sup>. Sin embargo, lo cierto es que para

<sup>2</sup> En un reciente informe oficial de la *British Psychological Society* elaborado por Bull, Baron, Gudjonsson, Hampson, Rippon y Vrij (2004), se presentan los resultados de diversas revisiones sobre la validez del polígrafo. Con el empleo de la Prueba de la Pregunta Control (CQT), el porcentaje de mentirosos identificados oscila, según la revisión considerada, entre el 83% y el 89%, y el porcentaje de persona veraces identificadas oscila entre el 53% y el 78%. Con el empleo de la Prueba del Conocimiento del Culpable (GKT), el polígrafo permite identificar prácticamente a todas las personas veraces (precisión del 98% y del 94% según la revisión considerada), pero posee una pobre capacidad para detectar a los mentirosos (42% y 76%) (Bull *et al.*, 2004). Entre los procedimientos verbales para evaluar la credibilidad destacan el Análisis de Contenido Basado en Criterios (CBCA) y la aproximación del Control de la Realidad (RM). Con el CBCA se pueden identificar correctamente un 73% de las declaraciones verdaderas y un 72% de las declaraciones falsas (Vrij, 2005). La precisión del RM es similar, alcanzando un nivel de discriminación del 72% al clasificar tanto declaraciones verdaderas como falsas (Masip, Sporer, Garrido y Herrero, 2005). Señalábamos en otro lugar (Masip, Garrido y Herrero, 2002b) que, a diferencia de los poligrafistas o los evaluadores que emplean el CBCA y el RM, los observadores de los experimentos llevados a cabo desde la aproximación no-verbal no están entrenados, por lo que la comparación es inadecuada. Sin embargo, según se señala más adelante en el texto, los incrementos obtenidos mediante el entrenamiento en indicadores no-verbales son muy limitados. Una metodología que permitió buenos resultados a partir del análisis del comportamiento no-verbal es la empleada por Vrij, Edward, Roberts y Bull (2000), si bien sus hallazgos deben ser replicados. Sobre este tema, véase Masip *et al.* (2002b).

<sup>3</sup> Más exactamente, la precisión fue menor cuando los observadores estuvieron expuestos al canal visual que cuando estuvieron expuestos a los canales auditivo y audiovisual; las comparaciones intraestudio (pero no las interestudio) mostraron que se detecta mejor a los emisores motivados que a los no-motivados; también sólo en las comparaciones intraestudio la precisión fue menor cuando el emisor había podido preparar el mensaje que cuando no lo había podido preparar; la exposición previa a la conducta habitual del emisor favoreció la detección; y las comparaciones intraestudio (no pudieron hacerse comparaciones interestudio por haberse variado este factor sólo en raras ocasiones) indicaron que durante una interacción emisor-receptor la detección es mayor que cuando el receptor observa un mensaje continuo e ininterrumpido enviado por el emisor (Bond y DePaulo, en prensa).

alguna de ellas (motivación y preparación) éste sólo apareció en las comparaciones intraestudio, pero no en las comparaciones interestudio. Y además, pese a la significación de algunas diferencias, prácticamente en todos los casos en que los autores señalan los índices de precisión éstos estuvieron por debajo del 60%. Por lo tanto la influencia de estas variables, pese a ser estadísticamente significativa, es realmente muy reducida en términos absolutos. Por su parte, el trabajo meta-analítico de Aamodt y Mitchell (2005) muestra que variables individuales tan importantes como la edad de los receptores, su sexo, su nivel educativo/capacidad cognitiva y los rasgos de extraversión y neuroticismo no se relacionan significativamente con la precisión de los juicios. Sólo la automonitorización parece tener una débil relación positiva con ella ( $r = .14$ ).

Estos resultados se refieren a la detección de mentiras y verdades (reflejan el porcentaje de clasificaciones correctas al considerar conjuntamente declaraciones verdaderas y falsas), pero ¿qué pasa específicamente con la detección de la mentira? La investigación muestra que las personas identificamos con mayor facilidad verdades que mentiras (Levine, Park y McCornack, 1999). Esto es así porque presentamos una tendencia a considerar que los demás dicen la verdad, lo cual incrementa nuestra precisión al juzgar verdades y la reduce al juzgar las mentiras (Levine *et al.*, 1999; Masip *et al.*, 2002b). Así, por ejemplo, en el meta-análisis de Bond y DePaulo (en prensa) se halló que el porcentaje medio de juicios de verdad fue del 55.0%, significativamente superior al 50% esperado por azar. Esto hizo que la precisión al juzgar declaraciones verdaderas se situara en el 60.3%, sensiblemente por encima de la precisión al juzgar declaraciones falsas, que alcanzó tan sólo la tasa del 48.7%.

Esta tendencia a juzgar las declaraciones como verdaderas puede deberse a varias razones (véase Levine *et al.*, 1999). Es posible que esté basada en un modo de procesamiento heurístico (Stiff, Kim y Ramesh, 1992), o en el propio funcionamiento de la mente, que en principio representaría como cierta toda aquella información entrante que comprende (Gilbert, Krull y Malone, 1990), o puede derivarse de la estrategia adaptativa de creer los mensajes que se reciben, ya que en la vida cotidiana

la mayor parte de ellos son ciertos (Anderson, Ansfield y DePaulo, 1999). Recientemente, sobre la base de dos estudios que muestran que cuanto más información se proporciona al receptor menor es el sesgo de veracidad, hemos propuesto que éste pudiera deberse a un artefacto experimental (Masip, Garrido y Herrero, 2005, en prensa). Ciertamente, en la investigación realizada hasta el momento los fragmentos de la conducta del emisor que se han utilizado como material estimular han sido muy breves. Esta brevedad limita la cantidad de información que el observador puede recibir del emisor, de forma que, a la hora de formular sus juicios, el observador se ve obligado a emplear un modo de procesamiento heurístico. Y se da la circunstancia de que en tareas de evaluación de la credibilidad los juicios heurísticos suelen ser juicios de verdad (véanse Gilbert *et al.*, 1990; Millar y Millar, 1997; Stiff *et al.*, 1992). Por lo tanto, el sesgo de veracidad hallado en la investigación pudiera deberse a la brevedad de las muestras conductuales empleadas. En consonancia con esta idea, hemos mostrado que el empleo de muestras de conducta más extensas e informativas reduce este sesgo (Masip, Garrido y Herrero, 2005, en prensa). Sin embargo, este hallazgo debe ser replicado por otros equipos de investigación, y quedan todavía algunas preguntas por responder (Masip, Garrido y Herrero, 2005, en prensa).

En cualquier caso, la tendencia a juzgar las declaraciones como verdaderas parece ser menor entre aquellos profesionales para quienes la detección de la mentira es más relevante que en otras personas (Bond y DePaulo, en prensa). Se ha llegado incluso a afirmar, sobre la base de los resultados empíricos, que en realidad tales profesionales presentan un sesgo opuesto que les lleva a considerar que las declaraciones son falsas (Meissner y Kassin, 2002), y que presentan una tendencia generalizada a cuestionar la veracidad de las comunicaciones emitidas por los demás<sup>4</sup> (Masip, Alonso, Garrido y Antón, 2005).

En resumen, la investigación revisada en este apartado muestra que: (a) la capacidad de los seres humanos para discriminar entre mensajes verdaderos y falsos es muy escasa; (b) esto es así incluso entre personas para quienes dicha discriminación tiene importancia profesional; (c) aunque hay algunas variables que afectan significati-

<sup>4</sup> Recientemente, Kassin, Meissner y Norwick (2005) han hallado que los policías tienden más que los no-policías a considerar verdaderas una serie de confesiones falsas de delitos, por lo que han modificado su visión inicial y sostienen que, más que un sesgo a considerar que las declaraciones son falsas, lo que presentan tales profesionales es un sesgo a considerar que los emisores de tales declaraciones son culpables.

vamente al nivel de aciertos, en términos absolutos las variaciones oscilan entre el 50% y el 60%, manteniéndose siempre por debajo de niveles de precisión aceptables; (d) la investigación muestra que las personas tendemos a prestar credibilidad a lo que otros nos dicen, por lo que detectamos más verdades que mentiras; sin embargo, hay indicios de que este resultado pudiera deberse al modo en que habitualmente se ha hecho la investigación; (e) por el contrario, los profesionales para quienes la evaluación de la credibilidad es importante muestran una tendencia a considerar que los mensajes son *falsos*.

### CONFIANZA: ¿SOMOS CONSCIENTES DE NUESTRA (IN)CAPACIDAD PARA DETECTAR MENTIRAS?

Una vez establecido que es difícil detectar mentiras sobre la base del comportamiento no-verbal pasamos a otra cuestión examinada por la investigación: ¿existe alguna relación entre la confianza depositada en nuestros juicios y nuestra precisión? DePaulo, Charlton, Cooper, Lindsay y Muhlenbruck (1997) hicieron un meta-análisis de la investigación sobre la confianza al hacer juicios de veracidad. Con la muestra de los 18 estudios relevantes que pudieron localizar, encontraron una correlación media prácticamente nula:  $r = .04$ . Aamodt y Mitchell (en prensa) han examinado la misma cuestión, añadiendo experimentos más recientes a los incluidos en el meta-análisis de DePaulo *et al.* (1997). La correlación promedio en 58 estudios hallada por Aamodt y Mitchell es virtualmente la misma:  $r = .05$ . En definitiva, las personas no tenemos conciencia de lo correctos o incorrectos que son nuestros juicios de credibilidad.

Otro hallazgo de interés relacionado con la confianza es la evidencia de que tendemos a sobreestimar nuestra capacidad de discriminar entre verdades y mentiras. DePaulo *et al.* (1997) compararon la confianza y la precisión en seis estudios en los que ambas variables se habían medido en una escala de 0 a 100 (o en los que las puntuaciones se podían transformar a dicha escala). Hallaron una

precisión media del 57.20% y una confianza media en sus juicios del 72.91%, claramente superior.

### INDICADORES: MÍRAME A LOS OJOS Y DIME LA VERDAD

Muchos libros populares sobre comunicación no-verbal presentan la detección de la mentira como una tarea sencilla: basta con examinar si el emisor muestra determinadas señales conductuales claramente visibles para determinar si está mintiendo o no. Por ejemplo, Lieberman (1998) y Pease (1981/1988) afirman que taparse la boca, tocarse la nariz, frotarse un ojo o el cuello o tirar del cuello de la camisa son indicación de que el interlocutor está mintiendo.

Asimismo, las personas tienen creencias muy claras sobre cuáles son los indicadores conductuales del engaño (véanse entre otras las revisiones de Strömwall *et al.*, 2004; Vrij, 2000). Por ejemplo, una creencia popular muy extendida (y que también se encuentra en el libro de Lieberman) es que los mentirosos apartan su mirada. En un reciente trabajo transcultural se ha hallado que este estereotipo tiene validez universal. Cuando se preguntó a personas de 58 países distintos “¿Cómo puedes saber si alguien está mintiendo?”, los habitantes de 51 de ellos mencionaron que las personas apartan la mirada al mentir (Global Deception Research Team, en prensa). En un segundo estudio se empleó un cuestionario con preguntas cerradas. Una de éstas aludía al contacto ocular. Las tres opciones de respuesta eran que la gente mira más a los ojos del interlocutor al mentir que al decir la verdad, que mira menos, y que mira en igual medida. En 61 de los 63 países estudiados los participantes escogieron la segunda de estas tres opciones con más frecuencia que ninguna de las otras dos (Global Deception Research Team, en prensa). ¿En qué medida son correctas tales creencias? ¿Existen indicadores claros del engaño? ¿Cuáles son?

En diversas revisiones se han comparado los resultados de los estudios centrados en los indicadores *reales* del

<sup>5</sup> Los indicadores *reales* del engaño se estudian comparando la medida en que diversas categorías conductuales (por ej., dirección de la mirada, tartamudeos, etc.) están presentes en comunicaciones verdaderas y falsas. Para examinar los indicadores *percibidos* del engaño la comparación se establece entre comunicaciones *juzgadas* verdaderas y comunicaciones *juzgadas* falsas por los observadores. Las *creencias o estereotipos* sobre los indicadores del engaño se estudian preguntando a las personas qué claves creen ellas que permiten diferenciar entre comunicaciones verdaderas o falsas. Como hemos visto al presentar los resultados del trabajo del Global Deception Research Team (en prensa), se pueden emplear preguntas abiertas o cerradas. Además, éstas pueden formularse en términos generales (“¿cómo puedes saber si alguien está mintiendo?”) o, como sucede en Masip, Garrido, Herrero, Antón y Alonso (en prensa), pueden referirse a un juicio o conjunto de juicios específicos (“¿en qué te has basado para concluir que esta persona estaba mintiendo/diciendo la verdad?”).



engaño (conductas que diferencian entre declaraciones verdaderas y falsas) con los de aquellos estudios que han examinado los indicadores *percibidos* o las *creencias* de la gente sobre los índices del engaño. Los indicadores percibidos son aquellos que las personas *utilizan realmente* para hacer sus juicios de credibilidad, y las creencias son los indicadores que las personas *dicen que* son útiles para discriminar entre verdades y mentiras<sup>5</sup> (Masip y Garrido, 2000, 2001). En general, las coincidencias entre las últimas dos categorías y la primera son muy escasas, reflejando que las personas tenemos un gran desconocimiento de las claves que realmente pueden discriminar entre comunicaciones verdaderas y falsas (Burgoon, Buller y Woodall, 1994; DePaulo, Stone y Lassiter, 1985; Vrij, 2000). Por ejemplo, Vrij (2000) observa que si bien la gente cree que, en comparación con quienes dicen la verdad, los mentirosos mueven más sus extremidades, desvían más la mirada, parpadean más, sonríen más, muestran más automanipulaciones y gestos ilustrativos, cambian con mayor frecuencia de postura y mueven más el tronco, los resultados de la investigación empírica muestran que, en realidad, los mentirosos mueven sus extremidades *menos* que los veraces, y que la relación entre las demás conductas y el engaño no es significativa. Otras creencias populares examinadas por Vrij, como que los mentirosos cometen más errores y presentan más vacilaciones al hablar, que hacen más pausas, etc., no han recibido apoyo claro de la investigación, puesto que se han hallado resultados contradictorios debido a que determinadas variables, como la complejidad cognitiva de la mentira, pueden mediar la expresión de las conductas relevantes. Hay dos creencias populares que, según Vrij, son acertadas: la de que al mentir se habla con un tono de voz algo más agudo y la de que las pausas al hablar son de mayor duración al mentir que al decir la verdad. En conclusión, la abrumadora mayoría de las creencias populares sobre los indicadores no-verbales del engaño son erróneas. Por desgracia, sucede lo mismo con las creencias que presentan profesionales tales como policías, jueces, etc., las cuales se solapan en gran medida con las del ciudadano medio (véase Strömwall *et al.*, 2004, para una discusión en profundidad).

Una posible explicación de esta falta de concordancia entre creencias y realidad nos la ofrece Kelley (1992) cuando hipotetiza que las nociones del sentido común probablemente sean menos válidas cuando se refieren al *micronivel* que cuando se refieren al *mesonivel*. En el *micronivel*, Kelley ubica los “acontecimientos que ocurren rápidamente ..., en escalas pequeñas de magnitud o masa (por ej., pequeñas contracciones de los músculos faciales o cambios en la fijación ocular), y a menudo de forma invisible...” (Kelley, 1992, p. 6). El *mesonivel* es el “nivel de la conducta individual molar...” (Kelley, 1992, p. 6), y comprende “consecuencias inmediatas y directas, periodos de tiempo de minutos a días ... Este nivel es el centro de atención en la vida diaria...” (Kelley, 1992, p. 6). Indudablemente, la identificación de claves discretas del engaño se inserta en el *micronivel* de Kelley.

Sea como fuere, la discrepancia entre los estereotipos populares y la realidad empírica puede dar cuenta del escaso valor de las claves conductuales para formular juicios *correctos* de mentira. Park, Levine, McCornack, Morrison y Ferrara (2002) preguntaron a un grupo de estudiantes que recordaran un caso en el que hubieran descubierto que otra persona les había mentido y que indicaran qué estrategias habían empleado en esa ocasión para descubrir el engaño. Los resultados muestran que los métodos más usados fueron la información de terceras personas, la evidencia material y la confesión del propio mentiroso. Prestar atención a las claves no-verbales y verbales estuvo entre las estrategias menos empleadas (2.1%). En definitiva, el papel de tales claves para formular juicios correctos de mentira es ínfimo<sup>6</sup>.

El trabajo de Vrij (2000) descrito anteriormente revisa sólo parte de la literatura. Con posterioridad al mismo, DePaulo, Lindsay, Malone, Muhlenbruck, Charlton y Cooper (2003) han publicado el trabajo meta-analítico más exhaustivo realizado hasta el momento sobre los indicadores no-verbales y verbales del engaño. Aunque no comparan tales indicadores con las creencias populares, sus resultados son del máximo interés, pues permiten aislar las claves que, potencialmente, pueden ser útiles para discriminar entre verdades y mentiras. DePaulo *et al.* examinaron un total de 116 informes de investigación en

<sup>6</sup> Park *et al.* (2002) interpretan los resultados como indicativos de que las personas no emplean indicadores verbales y no-verbales para hacer sus juicios de credibilidad. Sin embargo, al haber limitado los autores su exploración a mentiras que llegaron a descubrirse, sólo podemos concluir que tales claves tienen un efecto limitado sobre los juicios correctos de mentira. Es posible que esas claves se utilicen con frecuencia pero que sean muy poco discriminativas.

los que se explora la relación de 158 claves conductuales con el acto de mentir o decir la verdad. Los autores diferenciaron entre dos conjuntos de claves. Primero, aquellas que se habían examinado por lo menos en tres ocasiones distintas, habiendo podido calcular con precisión el tamaño del efecto para al menos dos de ellas. El tamaño del efecto es, en este caso, un índice de la relación entre la presencia/ausencia de la clave y si el emisor miente o dice la verdad. Sólo puede calcularse con exactitud si se proporciona la suficiente información en los informes de investigación originales, lo cual no sucedía en todos los examinados por DePaulo *et al.* El segundo conjunto de claves comprendía a todas las demás. Los cálculos referentes al primer conjunto son más válidos, dado el mayor número de muestras y la mayor precisión en los cálculos del tamaño del efecto.

Los autores hallaron que sólo 24 claves de las 88 del primer grupo diferenciaron entre declaraciones verdaderas y falsas. A éstas se añadieron 17 del segundo grupo. En conjunto,  $24 + 17 = 41$  claves de un total de las 158 examinadas; esto es el 26.0%. Si sólo consideramos las 24 claves significativas del primer grupo, cuyo cálculo presenta más garantías, el porcentaje es del 15.2%. En conclusión, a diferencia de lo que promulgan una serie de libros "de autoayuda" y de lo que sostiene la sabiduría popular, hay muy pocas diferencias entre la conducta de las personas cuando mienten y cuando dicen la verdad.

Con el fin de aislar los indicadores más válidos del engaño, DePaulo *et al.* (2003) se centraron en aquellos basados en un número de comparaciones superior a cinco y con un tamaño del efecto igual o superior a 0.20 en valores absolutos. Sólo hallaron 12 de tales indicadores, la mayoría de naturaleza verbal. La clave más discriminativa ( $d = -0.55$ ) parece ser la inmediatez verbal y vocal. Esto significa que al mentir las personas responden de manera menos directa, relevante y clara que al decir la verdad, y que además lo hacen de forma evasi-

va e impersonal (DePaulo *et al.*, 2003). Además, en comparación con las comunicaciones de quienes dicen la verdad, las comunicaciones de los mentirosos parecerán más ambivalentes y discrepantes (por ej., habrá falta de concordancia entre lo expresado a través de unos canales y otros) ( $d = 0.34$ ). Asimismo, las mentiras tendrán menos detalles ( $d = -0.30$ ), una estructura menos lógica ( $d = -0.25$ ) y un menor engranaje contextual ( $d = -0.21$ ) que las verdades. Éstos son tres criterios verbales del Análisis de Contenido Basado en Criterios o CBCA<sup>7</sup> (Garrido y Masip, 2000, 2004; Masip, Garrido y Herrero, 2003; Vrij, 2005). Las narraciones falsas también parecerán menos plausibles ( $d = -0.23$ ) y contendrán más afirmaciones negativas y quejas ( $d = 0.21$ ) que las verdaderas. El narrador parecerá inseguro y vacilante en su voz y en sus palabras ( $d = 0.30$ ), dará la impresión de estar más nervioso o tenso ( $d = 0.27$ ), su voz también sonará tensa ( $d = 0.26$ ) y de hecho su tono fundamental (frecuencia de la voz) será más agudo ( $d = 0.21$ ). Además, la implicación personal del narrador a nivel verbal y no-verbal será menor en declaraciones falsas que en declaraciones verdaderas ( $d = -0.21$ ). Es importante señalar que ninguna de las pintorescas claves antes mencionadas que describe Pease (1981/1988) se encuentra en esta lista basada en un riguroso meta-análisis de la investigación relevante, ni tampoco el contacto ocular<sup>8</sup>.

Es extremadamente importante tener en cuenta que estos resultados se derivaron de *todo* el conjunto de estudios y condiciones experimentales de los trabajos analizados por DePaulo *et al.* (2003). Pero se detectaron una serie de circunstancias que influyen sobre la utilidad de los indicadores para discriminar entre declaraciones verdaderas y falsas. Así, la motivación del emisor, el objeto que se persigue con el engaño (ocultar una transgresión vs. otros fines), la extensión de la respuesta (tiempo durante el que el emisor se expresa) y la preparación previa de la mentira influyeron sobre el significado y el

<sup>7</sup> La *estructura lógica* implica que los diversos detalles describen idéntico curso de sucesos, la declaración en su conjunto es coherente y lógica y sus partes "encajan". Por *engranaje contextual* se entiende que el acontecimiento descrito está inserto en un contexto espacio-temporal rico y complejo (véase Garrido y Masip, 2001).

<sup>8</sup> El tamaño del efecto para el contacto ocular fue  $d = 0.01$ , y para la conducta de desviar la mirada  $d = 0.02$ ; ambas  $ds$  fueron no-significativas. Las claves que arrojaron tamaños del efecto superiores a 0.20 en valores absolutos pero que se calcularon sobre la base de 5 o menos comparaciones (en realidad 3 a 5 comparaciones) fueron cooperatividad ( $d = -0.66$ ), admisión de falta de memoria ( $d = -0.42$ ), dilatación pupilar ( $d = 0.39$ ), duración del discurso ( $d = -0.35$ ), asociaciones externas relacionadas ( $d = 0.35$ ), inmediatez verbal ( $d = -0.31$ ), correcciones espontáneas ( $d = -0.29$ ), elevación de la barbilla ( $d = 0.25$ ), atribuciones sobre el estado mental del otro ( $d = 0.22$ ), repeticiones de palabras y frases ( $d = 0.21$ ) y autodesaprobación ( $d = 0.21$ ). Los valores positivos de  $d$  indican que la conducta se muestra más al mentir que al decir la verdad; los valores negativos tienen el significado opuesto.

poder discriminativo de diversos indicadores (DePaulo *et al.*, 2003; DePaulo y Morris, 2004). Por ejemplo, cuando la comunicación no estaba preparada de antemano la latencia de respuesta (tiempo transcurrido entre el final de la pregunta y el inicio de la respuesta del emisor) fue mayor al mentir que al decir la verdad, pero cuando la comunicación estaba preparada de antemano la latencia fue mayor al decir *la verdad* que al mentir. Asimismo, hubo varias claves (por ej., parpadeos) que discriminaron cuando se mentía sobre transgresiones pero que no discriminaron al mentir sobre otros temas (para una descripción completa de los efectos de las variables moderadoras sobre los indicadores, véanse DePaulo *et al.*, 2003; DePaulo y Morris, 2004). En resumen: (a) el significado de los mismos indicadores (por ej., latencia de la respuesta) puede cambiar según las circunstancias; (b) hay conductas (por ej., parpadeos) que discriminan significativamente en unas circunstancias pero no en otras; y (c) hay claves (por ej., parpadeos) que no discriminan en términos generales pero que sí lo hacen en circunstancias muy específicas, y viceversa. Así pues, al contrario de lo que se afirma en muchos libros de autoayuda, no sólo hay pocas claves del engaño, sino que éstas son muy específicas de cada situación. Como señala Kelley (1992), el sentido común es más sensible a los efectos principales que a las interacciones que la ciencia desvela, y además la ciencia descubre factores subyacentes que no están en el punto de vista del observador lego y que tienen gran influencia en los resultados.

### ENTRENAMIENTO: ¿EXISTE ALGUNA REMOTA ESPERANZA?

El panorama que se dibuja en las páginas anteriores es ciertamente desolador: los seres humanos somos pésimos detectores de mentiras, nuestra confianza no se relaciona con la precisión de nuestros juicios, tendemos a sobreestimar nuestra capacidad de detectar mentiras, nuestras creencias sobre los indicadores del engaño son erróneas y utilizamos claves equivocadas al hacer tales juicios. ¿Existe alguna esperanza de aprender a hacerlo bien?

Se han realizado muchos intentos de entrenar a las personas para detectar el engaño (véanse las revisiones de Bull, 2004; Frank y Feeley, 2003; Vrij, 2000). Vrij observa que se han utilizado tres tipos de entrenamiento. Uno consiste en proporcionar a los sujetos *retroalimenta-*

*ción* sobre sus resultados, de forma que puedan aprender de sus errores y sus aciertos al ir efectuando los juicios de credibilidad. Otro tipo de entrenamiento se basa en una *estrategia informacional*, consistente en indicar a los observadores cuál es la verdadera relación entre determinados indicadores y el engaño. Un tercer tipo de entrenamiento se basa en una *estrategia atencional*, en que se focaliza la atención de los observadores sobre determinadas claves reveladoras (sin explicitar necesariamente su significado), o bien sobre aquellos canales más transparentes (por ej., el canal auditivo). Según Vrij, con independencia del método empleado, en general los observadores han logrado incrementar su nivel de aciertos en la condición de entrenamiento. Pero el autor también indica que tales incrementos han sido muy pobres: precisión media del 54% en los grupos no-entrenados vs. del 57% en los grupos entrenados.

En un trabajo posterior al de Vrij (2000) y más sistemático que éste, Frank y Feeley (2003) meta-analizan la investigación realizada hasta el momento sobre el entrenamiento no-verbal en detección de la mentira. Su trabajo considera 20 comparaciones efectuadas en 11 trabajos publicados, con un total de 1072 observadores en los grupos de entrenamiento y 1161 en los grupos control. Encuentran que el incremento de precisión debido al entrenamiento es estadísticamente significativo, pero muy pequeño: se informa de un nivel medio de aciertos del 54% en los grupos no-entrenados y del 58% en los grupos entrenados; nótese que los valores son casi idénticos a los hallados por Vrij (2000). Los autores argumentan que la escasa calidad de los programas de entrenamiento empleados puede estar detrás de tan pobre incremento. Sin embargo, aunque es cierto que los programas empleados presentan una serie de limitaciones, un problema más fundamental atañe a la escasa relación, antes señalada, entre indicadores conductuales y el engaño, así como el relativismo de esta relación en función de diversas circunstancias (DePaulo *et al.*, 2004). Esto puede afectar negativamente a la eficacia de las tres modalidades de entrenamiento identificadas por Vrij (2000). Así, lo que se pueda aprender mediante la *retroalimentación* en un programa del primer tipo será confuso, relativo y de escaso valor. En el caso de una *estrategia informacional*, poca será la información consistente y válida a nivel transituacional que pueda proporcionarse a los observadores. Por último, el empleo de una *estrategia atencional* también

presenta problemas. Si se orienta a los observadores a que focalicen su atención sobre determinadas *claves discretas*, éstas tendrán necesariamente una validez limitada y dependiente de las circunstancias. Y si lo que se pretende es focalizar la atención de los observadores sobre los *canales* auditivo y audiovisual, significativamente más transparentes que el canal meramente visual en el meta-análisis de Bond y DePaulo (en prensa), antes se debe tener en cuenta que, en las comparaciones interestudio (Bond y DePaulo no presentan los índices concretos de precisión en las comparaciones intraestudio), los niveles promedios de precisión alcanzados ante tales canales fueron del 53.7% (canal auditivo) y del 53.9% (canal audiovisual), frente al 50.2% del canal visual. Recuérdese que el nivel de aciertos por azar está en el 50%, y que la precisión total corresponde al 100%. Poca es, en consecuencia, la precisión final que podrán alcanzar los observadores al pedirles que presten atención a los canales auditivo o audiovisual.

Sobre la base de un análisis parcial de la investigación relevante, Meissner y Kassin (2002) sugieren que, más que incrementar la precisión, lo que hacen los programas de entrenamiento es incrementar la tendencia de los observadores a decir que los mensajes son falsos. De forma consistente con tales apreciaciones, en el meta-análisis más amplio de Frank y Feeley (2003) el incremento debido al entrenamiento fue nulo al juzgar verdades (precisión del 58% en los grupos no-entrenados vs. 56% en los entrenados), pero sustancial al juzgar mentiras (49% vs. 55%). Este efecto no debe sorprender. Aunque Vrij (2000) identificara las tres aproximaciones descritas anteriormente, en realidad la mayoría de los programas de entrenamiento se han basado en la estrategia de informar a los observadores sobre la supuesta relación entre ciertas claves conductuales y el engaño. Normalmente tales entrenamientos se centran específicamente sobre los indicadores *de la mentira*, y no sobre los indicadores de la verdad. Se señalan ciertas conductas, se dice que suelen aparecer con más frecuencia al mentir que al decir la verdad, y se invita a los observadores a que traten de identificarlas en los vídeos experimentales para determinar si los emisores *están mintiendo* (y no para diferenciar si los emisores *mienten o dicen la verdad*). Pero el que determinadas claves aparezcan con *mayor frecuencia* al mentir que al decir la verdad, no significa que aparezcan *exclusivamente* cuando se mien-

te. De modo que los observadores buscarán activamente esas claves indicadoras de engaño, y en cuanto perciban su más mínimo atisbo resolverán de inmediato y con firmeza que el emisor está mintiendo. Ésta puede ser la razón de que el entrenamiento incrementa sólo la frecuencia de juicios de mentira, pero no la precisión al juzgar verdades. Probablemente, un entrenamiento focalizado sobre las *claves de la verdad*, o bien un entrenamiento más equilibrado en el que se presenten, con idéntico énfasis, los indicadores *de la verdad* y los *de la mentira* (sus opuestos), y en el que la tarea no consista en *detectar mentiras*, sino en *discriminar entre declaraciones verdaderas y falsas*, tendría efectos muy distintos. Nuestra investigación más reciente está explorando esta posibilidad.

## CONCLUSIONES

La sabiduría popular sostiene que “es más fácil pillar a un mentiroso que a un cojo”. La mayoría de personas muestra gran confianza en sus juicios de veracidad. Existen además claros estereotipos populares sobre el comportamiento de las personas al mentir. Se encuentra asimismo en el mercado un sinnúmero de libros “de autoayuda”, que cuentan con gran aceptación popular, en los que se presenta la detección de la mentira a partir del comportamiento no-verbal como una tarea sencilla de aprender, y en los que se ofrecen extensas relaciones de supuestos indicadores del engaño de validez universal.

Frente a las creencias populares y a las afirmaciones de los libros “de autoayuda”, se han presentado en estas páginas los resultados de varias décadas de rigurosa investigación realizada por psicólogos y comunicólogos. Es importante que el lector tenga en cuenta que la mayor parte de los hallazgos descritos en el presente trabajo proviene de estudios meta-analíticos muy abarcativos, por lo que las muestras son extremadamente amplias y heterogéneas (y, por ende, representativas), y los resultados reflejan fielmente los hallazgos globales de virtualmente toda la investigación realizada. Tales resultados se oponen frontalmente a las creencias populares y a lo que se afirma en la mayoría de los libros “de autoayuda”. Así, se concluye lo siguiente: (a) la capacidad del ser humano para discriminar entre verdades y mentiras es extremadamente limitada; esto es así incluso en grupos profesionales para quienes la detección del engaño es una tarea importante en su trabajo; (b) las personas

no tenemos conciencia de lo correctos o incorrectos que son nuestros juicios de credibilidad; (c) tendemos a sobreestimar nuestra capacidad de identificar verdades y mentiras; (d) utilizamos claves equivocadas al hacer juicios de credibilidad; (e) las creencias populares sobre los indicadores del engaño son erróneas; (f) las creencias de los profesionales para quienes la detección del engaño es una tarea importante son también erróneas y similares a las de las otras personas; (g) no se ha demostrado que los indicadores conductuales que se mencionan en la mayoría de los libros "de autoayuda" permitan una adecuada discriminación entre verdades y mentiras; (h) existen muy pocas conductas que realmente permitan diferenciar entre verdades y mentiras; (i) al contrario de lo que se da a entender en muchos libros "de autoayuda" y de lo que sostiene la sabiduría popular, el significado y el poder de discriminación de las claves conductuales dependen de una serie de variables situacionales; (j) también al contrario de lo que afirman determinados libros dirigidos al gran público, aprender a discriminar entre verdades y mentiras es extremadamente difícil, como muestra la limitada eficacia de distintos programas de entrenamiento; y (k) en lugar de incrementar la precisión global, los entrenamientos al uso aumentan el sesgo a decir que las declaraciones son falsas.

En ocasiones, determinados colectivos profesionales cuya labor les exige evaluar la credibilidad se dejan llevar por sus creencias ingenuas. Otras veces, en un loable afán de aprender y capacitarse profesionalmente, buscan información en determinados libros, en muchas ocasiones aparentemente escritos por reputados profesionales de la psicología, pero que de hecho son obra de autores poco cualificados que sólo ofrecen ingenuos consejos de nulo valor científico. En otras ocasiones van más allá y asisten a cursillos o seminarios; pero a menudo éstos son impartidos por personas ajenas a los campos de la psicología o de la comunicación, o por compañeros más experimentados que, en muchos casos con la mejor de las intenciones, se limitan a transmitir sus intuiciones y creencias de sentido común, desvinculadas del avance científico en el campo de conocimiento relevante. En determinados ámbitos, las consecuencias de un juicio erróneo de la credibilidad pueden ser devastadoras (condena de un inocente; limitación del acceso a determinado empleo o su pérdida; etc.), por lo que es necesario que quienes deban hacer tales juicios reci-

ban la información más rigurosa y actualizada en el área de la detección del engaño. Los psicólogos están entre ellos, pero tienen además la importante responsabilidad adicional de asesorar a otros profesionales (y a legos) sobre la verdadera relación entre las claves conductuales y el engaño. En este sentido, quisiera haber podido ofrecer una lista clara de indicadores conductuales específicos, claramente perceptibles, y carentes de ambigüedad que fueran indicadores incuestionables de la mentira. Esto es lo que hacen los libros "de autoayuda", pero, por desgracia, la realidad es mucho más compleja. Ésta es la lección que conviene aprender.

**Agradecimientos:** El autor desea expresar su agradecimiento a Eugenio Garrido, Nuria Hernández y Roberto Vivero por su amabilidad al acceder a leer versiones anteriores de este trabajo y formular comentarios y sugerencias de gran ayuda.

#### REFERENCIAS

- Aamodt, M. y Mitchell, H. (en prensa). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner*.
- Abozzi, P. (1997). *La interpretación de los gestos*. Barcelona: Martínez Roca. (Publicado originalmente en italiano en 1996 por L'Airone Editrice, Roma, Italia).
- Anderson, D. E., Ansfield, M. E. y DePaulo, B. M. (1999). Love's best habit. Deception in the context of relationships. En P. Philippot, R. S. Feldman y E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 372-409). Nueva York: Cambridge University Press.
- Blair, J. P. y Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science and Management*, 6, 77-83.
- Bond, C. F., Jr. y DePaulo, B. M. (en prensa). Accuracy of deception judgments. *Personality and Social Psychology Review*.
- Bull, R. (2004). Training to detect deception from behavioral cues: Attempts and problems. En P.-A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 251-268). Cambridge: Cambridge University Press.
- Bull, R., Baron, H., Gudjonsson, G., Hampson, S., Rippon, G. y Vrij, A. (2004). *A review of the current scientific status and fields of application of polygraph*



- hic deception detection. Londres: British Psychological Society.
- Buller, D. B. y Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6, 203-242.
- Burgoon, J. K., Buller, D. B. y Woodall, W. G. (1994). *Nonverbal communication*. Columbus, OH: Greyden Press.
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J. y Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, 1(4), 346-357.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. y Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118.
- DePaulo, B. M. y Morris, W. (2004). Cues to deception and indirect lie detection. En P.-A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 15-40). Cambridge: Cambridge University Press.
- DePaulo, B. M. y Pfeiffer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 249-267.
- DePaulo, B. M., Stone, J. I. y Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323-370). Nueva York: McGraw-Hill.
- Drizin, S. A. y Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review*, 82, 891-1004.
- Frank, M. G. y Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31(1), 58-75.
- Garrido, E., Herrero, C. y Masip, J. (2004). Psicología jurídica y sentido común: Construcción social. *Revista de Psicología General y Aplicada*, 57, 395-406.
- Garrido, E. y Masip, J. (2004). La evaluación del abuso sexual infantil. En las *Actas del I Congreso de Psicología Jurídica y Forense en Red* (CD-Rom). Madrid: Colegio Oficial de Psicólogos de Madrid.
- Garrido, E., Masip, J. y Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology*, 39, 254-275.
- Gilbert, D. T., Krull, D. S. y Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601-613.
- Global Deception Research Team (en prensa). A world of lies. *Journal of Cross-Cultural Psychology*.
- Inbau, F. E., Reid, J. E., Buckley, J. P. y Jayne, B. C. (2001). *Criminal interrogation and confessions* (4<sup>ed.</sup>). Gaithersburg, MD: Aspen.
- Kassin, S. M. (2004). True or false: "I'd know a false confession if I saw one". En P.-A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 172-194). Cambridge: Cambridge University Press.
- Kassin, S. M. (2005). On the psychology of confessions. *American Psychologist*, 60, 215-228.
- Kassin, S. M. y Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, 23, 499-516.
- Kassin, S. M. y Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest*, 5, 33-67.
- Kassin, S. M., Meissner, C. A. y Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior*, 29, 211-227.
- Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology*, 43, 1-23.
- Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, 30, 209-216.
- Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., Bar, R., Pierro, A. y Mannetti, L. (2005). Says who? Epistemic authority effects in social judgment. *Advances in Experimental Social Psychology*, 37, 345-392.
- Lazarsfeld, P. F. (1949). The American Soldier –an expository review. *Public Opinion Quarterly*, 13, 377-404.
- Levine, T. R., Park, H. S. y McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs*, 66, 125-144.
- Lieberman, D. J. (1998). *Never be lied to again*. Nueva York: St. Martin's Press.
- Mann, S., Vrij, A. y Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal*

- of *Applied Psychology*, 89, 137-149.
- Masip, J., Alonso, H., Garrido, E. y Antón, C. (2005). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, 35(5), 1046-1066.
- Masip, J. y Garrido, E. (2000). La evaluación de la credibilidad del testimonio en contextos judiciales a partir de indicadores conductuales. *Anuario de Psicología Jurídica*, 10, 93-131.
- Masip, J. y Garrido, E. (2001). La evaluación psicológica de la credibilidad del testimonio. En F. Jiménez (Ed.), *Evaluación psicológica forense 1: Fuentes de información, abusos sexuales, testimonio, peligrosidad y reincidencia* (pp. 141-204). Salamanca: Amarú.
- Masip, J., Garrido, E. y Herrero, C. (2002a). La detección de la mentira mediante la técnica SCAN. *Revista de Psicopatología Clínica, Legal y Forense*, 2, 39-62.
- Masip, J., Garrido, E. y Herrero, C. (2002b). La detección del engaño sobre la base de sus correlatos conductuales: La precisión de los juicios. *Anuario de Psicología Jurídica*, 12, 37-55.
- Masip, J., Garrido, E. y Herrero, C. (2003). El Análisis de Contenido Basado en Criterios (CBCA). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 15(1), 75-92.
- Masip, J., Garrido, E. y Herrero, C. (2004). La detección de la mentira mediante la medida de la tensión en la voz: Una revisión crítica. *Estudios de Psicología*, 25, 13-30.
- Masip, J., Garrido, E. y Herrero, C. (2005). *Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias*. Enviado para su publicación..
- Masip, J., Garrido, E. y Herrero, C. (en prensa). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *International Journal of Psychology*.
- Masip, J., Garrido, E., Herrero, C., Antón, C. y Alonso, H. (en prensa). Officers as lie detectors. Guilty before charged. En D. Chadee y J. Young (Eds.), *Current themes and theories in social psychology*. St. Augustine: The University of the West Indies Press.
- Masip, J., Sporer, S. L., Garrido, E. y Herrero, C. (2005). The detection of deception with the Reality Monitoring approach: A review of the empirical evidence. *Psychology, Crime, & Law*, 11(1), 99-122.
- Meissner, C. A. y Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, 26, 469-480.
- Millar, M. G. y Millar, K. (1997). The effects of cognitive capacity and suspicion on truth bias. *Communication Research*, 24(5), 556-570.
- Miller, G. R. y Stiff, J. B. (1993). *Deceptive communication*. Newbury Park, CA: Sage.
- Myers, D. G. (2000). *Psicología social* (6ª ed.). Santa Fe de Bogotá: McGraw-Hill. (Publicado originalmente en inglés en 1999 por McGraw-Hill, Nueva York, EE.UU).
- National Research Council. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press.
- Park, H. S., Levine, T. R., McCornack, S., Morrison, K. y Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, 69, 144-157.
- Pease, A. (1988). *El lenguaje del cuerpo. Cómo leer el pensamiento de los otros a través de sus gestos*. Barcelona: Paidós. (Publicado originalmente en inglés en 1982 por Sheldon Press, Londres, Reino Unido).
- Schlesinger, A., Jr. (1949). The statistical soldier. *Partisan Review*, 16, 852-856.
- Stiff, J. B., Kim, H. J. y Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research*, 19(3), 326-345.
- Strömwall, L. A., Granhag, P.-A. y Hartwig, M. (2004). Practitioners' beliefs about deception. En P.-A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229-250). Cambridge: Cambridge University Press.
- Teigen, K. H. (1986). Old truths or fresh insights? A study of students' evaluations of proverbs. *British Journal of Social Psychology*, 25, 43-50.
- Vrij, A. (2000). *Detecting lies and deceit. The psychology of lying and the implications for professional practice*. Chichester: Wiley.
- Vrij, A. (2005). Criteria-Based Content Analysis. A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11(1), 3-41.
- Vrij, A., Edward, K., Roberts, K. y Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24(4), 239-263.

# EL ANÁLISIS DE CONTENIDO BASADO EN CRITERIOS (CBCA) EN LA EVALUACIÓN DE LA CREDIBILIDAD DEL TESTIMONIO

## CRITERIA-BASED CONTENT ANALYSIS (CBCA) IN STATEMENT CREDIBILITY ASSESSMENT

Verónica Godoy-Cervera, Lorenzo Higuera  
Universidad de Granada

*Este trabajo está dirigido por un lado a la descripción del Análisis de Contenido Basado en Criterios (CBCA), procedimiento enfocado originalmente a la evaluación de la credibilidad del testimonio de niños, y por otro, a la presentación de algunas investigaciones realizadas en torno a este tema. En los últimos años, sin embargo, ha habido un creciente interés por generalizar la aplicación de este procedimiento a adultos, por lo que otro de nuestros objetivos será comentar los estudios encaminados a emplear esta prueba en la evaluación de la credibilidad de las declaraciones de adultos. Finalmente comentaremos algunos inconvenientes de esta técnica y futuras líneas de investigación.*

*This work is directed on the one hand to the description of the Criterial Based Content Analysis (CBCA), a procedure focused originally on the evaluation of children's testimony credibility, and by other one, to present several investigations done regarding this topic. Over the last years, nevertheless, there has been an increasingly interest about the application of this procedure on adults, therefore, another purpose was to bring up studies focussed to the use of statement credibility analysis technique on adults. Finally, disadvantages and future research on the field on the use of this technique are discussed.*

**N**o es poco común que en las investigaciones policiales se cuente únicamente con la declaración de la víctima y la declaración del acusado como únicas pruebas disponibles del delito. Ante esta necesidad, varios investigadores se han centrado en el desarrollo de métodos sistemáticos que ayuden a identificar informes honestos de aquellos informes que han sido fabricados. Vrij (2000) y Vrij, Edward y Bull (2001) han clasificado estos procedimientos en tres grupos. El primero se centra en el registro y análisis de la actividad psicofisiológica de la persona que miente. El segundo se encarga de examinar la conducta no verbal del sujeto (Vrij, Edward, Roberts y Bull, 2000). El tercer grupo, y sobre el cual nos vamos a centrar, se encarga del estudio del contenido de la declaración del testigo (Masip, Sporer, Garrido y Herrero, 2005; Ruby y Brigham, 1997).

### DESARROLLO DEL ANÁLISIS DE CONTENIDO BASADO EN CRITERIOS (CBCA)

El *Statement Validity Assessment* (SVA) (*Evaluación de la Validez de la Declaración*) es la técnica más empleada para evaluar la veracidad de las declaraciones verbales (Vrij, 2000). El SVA se desarrolló en Alemania y se sustentó en la experiencia clínica de diversos psicólogos.

*Correspondencia:* Verónica Godoy Cervera. Facultad de Psicología. Universidad de Granada 18071. Granada España. E-mail: vgodoy@ugr.es

Alrededor de 1950, Udo Undeutsch realizó la primera descripción del SVA (ver Undeutsch, 1989) y posteriormente fue modificada hasta su forma actual por Steller y Köhnken (1989) y Raskin y Esplin (1991). En un principio, el SVA se desarrolló para valorar las declaraciones verbales de niños que habían sido víctimas de abuso sexual. Sin embargo, en años recientes se ha intentado validar y generalizar la aplicación de este instrumento a adultos (Vrij et al., 2001; Vrij et al., 2000). A pesar de que es un instrumento ampliamente utilizado en el ámbito forense como prueba psicológica no se le debe considerar un *test* o una escala estandarizada, sino un método semi-estandarizado para la evaluación de la credibilidad de las declaraciones (Steller, 1989). El desarrollo del SVA está fundamentado en lo que Steller (1989) ha denominado la hipótesis de Undeutsch. De acuerdo a esta hipótesis, un testimonio basado en una experiencia real difiere en cuanto a su calidad y contenido de un testimonio basado en un acontecimiento imaginado.

El *Criteria-Based Content Análisis* (CBCA) (*Análisis de Contenido Basado en Criterios*) es el componente principal del SVA lo que lo ha llevado a ser el elemento más frecuentemente estudiado por los investigadores (Ruby y Brigham, 1997) y sobre el cual se va a centrar este trabajo. El SVA está formado por tres componentes mutuamente dependientes: a) una entrevista estructurada con la víctima, b) el CBCA que evalúa el contenido de la de-

claración de la persona, y c) la integración del CBCA con la información derivada de un set de preguntas denominado *Lista de Validez*, el cual combina la información extraída del análisis del contenido de la declaración con otra información relevante del caso y con la información obtenida a partir de la exploración de la entrevista o entrevistas previamente realizadas (Horowitz, 1991).

La entrevista debe preceder a la aplicación de los criterios del CBCA. El objetivo primordial es obtener material sobre el cual aplicar dichos criterios. Es importante que el entrevistador esté familiarizado con el contenido de los criterios ya que de ello dependerá que la entrevista se desarrolle de forma adecuada. Así mismo, debe intentar obtener la mayor cantidad posible de información empleando una entrevista diseñada de forma que maximice la cantidad de información aportada por el testigo y minimice cualquier tipo de contaminación generada ya sea por el entrevistador o por cualquier otro adulto (Raskin y Esplin, 1991).

El CBCA se aplica al contenido de la declaración y su propósito es determinar si su calidad y sus contenidos específicos son indicativos de una narración generada a partir de registros de memoria o si son producto de la invención, la fantasía o la influencia de otra persona. Cualquier análisis realizado empleando el CBCA se ve influenciado por las características de la entrevista y por lo que el sujeto ha experimentado o no. A lo largo de este proceso, es importante que el entrevistador tome en cuenta la edad de la persona, su experiencia y el nivel de sus habilidades cognitivas (Raskin y Esplin, 1991). Una de las mayores limitaciones del CBCA es la dificultad que presenta al ser aplicado a situaciones en las que el testigo tiene información a partir de la cual puede inventar una acusación que incorpore algunos de los criterios. Por ejemplo, un niño que ha sido abusado sexualmente con anterioridad puede suministrar un testimonio falso pero que parezca convincente derivado de registros de memoria originados de otras experiencias. Este aspecto debe ser considerado cuando se lleve a cabo la revisión del caso a través de la *Lista de Validez* (Raskin y Esplin, 1991). El contenido verbal de la declaración es analizado mediante la aplicación de una serie de 19 criterios (véase Tabla 1), los cuales están organizados dentro de cinco grandes categorías y tienen como finalidad diferenciar entre declaraciones verdaderas y declaraciones fabricadas. Se parte de la idea de que un testimonio veraz contiene un mayor número de criterios

(para una descripción detallada de estos criterios véase Steller y Köhnken, 1989).

El análisis de la entrevista a través de los 19 criterios del CBCA se realiza otorgando puntuaciones numéricas a cada uno de los criterios. Según indica Steller (1989), se asigna 2, 1 ó 0 puntos en función de si el criterio se encuentra fuertemente presente, presente o ausente en la declaración. Otros autores (Lamb, Sternberg, Esplin, Hershkowitz, Orbach y Hovav, 1997) proponen se puntúe 1 ó 0 si el criterio está presente o ausente en la declaración. Ni la entrevista ni los resultados obtenidos a partir de los criterios son completamente válidos hasta que hayan sido puestos en contexto por medio de la *Lista de Validez*, la cual, está compuesta por cuatro categorías generales de información (Steller y Köhnken, 1989):

a) *Características psicológicas*. En esta categoría es importante evaluar la adecuación del lenguaje y el afecto y la susceptibilidad a la sugestión.

**TABLA 1**  
**EL ANÁLISIS DE CONTENIDO BASADO EN CRITERIOS (CBCA).**  
**MODIFICADO DE STELLER Y KÖHNKEN (1989)**

Características Generales
1. Estructura lógica. 2. Elaboración desestructurada. 3. Cantidad de detalles.
Contenidos específicos
4. Engranaje contextual. 5. Descripción de interacciones. 6. Reproducción de la conversación. 7. Complicaciones inesperadas durante el incidente.
Peculiaridades del contenido
8. Detalles inusuales. 9. Detalles superfluos. 10. Incomprensión de detalles relatados con precisión. 11. Asociaciones externas relacionadas. 12. Alusiones al estado mental subjetivo. 13. Atribución del estado mental del autor del delito.
Contenidos referentes a la motivación
14. Correcciones espontáneas. 15. Admitir fallos de memoria. 16. Plantear dudas sobre el testimonio. 17. Auto-desaprobación. 18. Perdón del autor de delito.
Elementos específicos de la ofensa
19. Detalles específicos de la ofensa.



- b) *Características de la entrevista.* El evaluador deberá realizar un análisis sobre la calidad de la entrevista valorando el tipo de preguntas formuladas (preguntas sugerentes, directivas o coactivas) y la adecuación global de la misma.
- c) *Motivación para realizar acusaciones falsas.* Esta categoría pretende descartar aquellos aspectos de índole motivacional que pudieran estar influyendo para que la persona proporcione una declaración falsa. No hay que olvidar también que el menor podría estar presionado por una tercera persona para falsear su testimonio. Un aspecto importante de esta categoría es hacer una valoración del contexto en el que se genera el informe.
- d) *Aspectos relacionados con la investigación.* Este apartado está diseñado con el fin de valorar la consistencia entre las declaraciones e investigaciones previas y partes médicos.

El evaluador deberá analizar la información relacionada con las cuatro categorías antes mencionadas y en función de ello determinar si dicha información apoya el testimonio de la persona. Así mismo, la *Lista de Validez* tiene por objetivo valorar varias hipótesis explicativas examinando toda la información del caso con la que se cuenta. Raskin y Esplin (1991) plantean que son cinco las hipótesis que deben ser comprobadas por el evaluador:

- a) La declaración es válida, pero el menor ha reemplazado la identidad del agresor por la de una persona distinta.
  - b) La declaración es válida, pero el menor ha sido influenciado o ha inventado información adicional que no es verdadera.
  - c) El menor ha sido presionado por una tercera persona para que formule una versión falsa de los hechos.
  - d) Por intereses personales o para ayudar a terceras personas el menor ha presentado una declaración falsa.
- 3) A consecuencia de problemas psicológicos, el menor ha fantaseado o inventado su declaración.

Es importante resaltar que el propósito del SVA es realizar una evaluación de la credibilidad del contenido de la declaración, no realizar una valoración sobre la credibilidad de la persona en sí (Steller y Köhnken, 1989).

Una de las grandes limitaciones del CBCA es que hasta ahora no se ha fijado una regla de decisión que nos ayude a establecer cuántos criterios determinan que una declaración sea clasificada como creíble o no creíble. Menos aún se ha precisado el peso que cada criterio de-

be recibir. Alonso-Quecuty (1999) plantea que el peso de cada criterio debe ser asignado tomando en cuenta diversos factores, tales como: el número de entrevistas previas por las cuales ha atravesado el niño, la complejidad del incidente, la edad del menor y el paso del tiempo. Una vez aplicados los criterios del CBCA y la *Lista de Validez*, el resultado final del análisis nos permite clasificar cualitativamente la declaración según cinco categorías (Alonso-Quecuty, 1999; Steller, 1989):

- Creíble.
- Probablemente creíble.
- Indeterminado.
- Probablemente increíble.
- Increíble.

#### INVESTIGACIONES REALIZADAS EN TORNO AL CBCA

En años recientes, los estudios sobre el CBCA han incrementado debido a su alta incidencia en el contexto judicial. Las líneas de investigación han seguido dos vertientes: 1) aquellas que emplean casos reales de menores que presuntamente han sido abusados sexualmente y en las que se utilizan otros elementos del caso como medidas de veracidad; y 2) estudios experimentales en los que se induce a los sujetos a manipular su declaración proporcionando ya sea un testimonio verdadero o uno falso (Ruby y Brigham, 1997).

#### *Estudios realizados con niños*

Debido a que el CBCA fue diseñado para ser aplicado a las declaraciones de menores presuntas víctimas de abuso sexual, la mayoría de las investigaciones publicadas han trabajado con muestras de estas características. A continuación se exponen brevemente algunos estudios realizados con menores.

Una investigación importante por cuanto trabaja con niños que presuntamente habían sufrido abuso sexual la realizaron Lamb et al. (1997). Su muestra estuvo compuesta por 98 niños israelíes (28 niños y 70 niñas), con edades de entre 4 y 13 años (media 8,72). Como medida de la veracidad de la declaración emplearon otros elementos del caso como: evidencia material o física, la declaración del acusado, etc. Tal y como habían predicho, hubo una mayor presencia de los criterios del CBCA en los relatos creíbles (media 6,74) en comparación con los relatos no creíbles (media 4,85). Sin embargo, los autores afirman que las diferencias encontradas no fueron tan significativas como lo fueron en estudios anteriores.



Más recientemente, Santtila, Roppola, Runtti y Niem (2000) analizaron el efecto que la edad, la habilidad verbal (medida con la escala verbal del WISC-R) y el estilo emocional del entrevistador tenían sobre la presencia de los criterios del CBCA en las declaraciones de 68 niños pertenecientes a tres grupos de edad diferente: 7-8, 10-11 y 13-14 años. En el experimento, se le pidió a cada niño que narrara dos experiencias personales, una real y una falsa. Los resultados mostraron una tasa de clasificación correcta del 66%. También comprobaron que la edad y la habilidad verbal del menor, así como la conducta del entrevistador influían en el número de criterios del CBCA que se presentaban en las declaraciones, independientemente de que éstas fueran verdaderas o falsas. Así mismo, hallaron que se presentaban diferentes criterios según el rango de edad en el que se encontraba el menor. Los autores sugieren integrar la información del CBCA con la lista de validez, y concluyen que las decisiones judiciales no deberían recaer exclusivamente en los resultados aportados por el CBCA-SVA.

Los estudios también se han centrado en analizar la influencia que la familiaridad del suceso a recordar tiene sobre la presencia de los criterios del CBCA. Por ejemplo, Pezdek et al. (2004) realizaron un experimento con 114 niños y se plantearon como hipótesis que las descripciones de eventos familiares tenían mayor probabilidad de ser catalogadas como veraces que las descripciones de eventos no familiares. Los resultados apuntaron a una mayor presencia de criterios del CBCA en las narraciones de sucesos familiares que en las narraciones de sucesos no familiares.

En esta misma línea, Blandon-Gitlin, Pezdek, Rogers y Brodie (2005) empleando una muestra de 94 niños analizaron la interacción entre la familiaridad del suceso y la veracidad del mismo. En el estudio se encontró que las puntuaciones obtenidas a través del CBCA fueron más fuertemente influidas por la familiaridad del suceso que por la veracidad del mismo. En ambos estudios, los autores concluyen sugiriendo que el CBCA, en su forma actual, es de limitada utilidad como herramienta para evaluar la credibilidad del testimonio en menores.

### **Estudios realizados con adultos**

Debido a los buenos resultados obtenidos con niños, la aplicación del CBCA ha intentado generalizarse a adultos. Sin embargo las investigaciones no son tan abun-

dantes como lo son en el caso de menores. A continuación comentaremos algunas de ellas.

Algunas investigaciones se han centrado en analizar qué criterios son los que se encuentran presentes en mayor medida en las declaraciones veraces y que por consiguiente serían los más sensibles en la discriminación de testimonios veraces y falsos. Por ejemplo, en un meta-análisis Ruby y Brigham (1997) encontraron que los criterios que más se presentaban en las declaraciones veraces eran el 1, 2, 3, 4, 5, 6, 7, 8, 12, 14, y 15. Por otro lado, Köhnken, Schimossek, Aschermann y Höfer (1995) hallaron que los testimonios veraces incluían un mayor número de detalles (criterio 3), eran desestructurados (criterio 4) y las personas tendían a admitir en mayor medida falta de memoria (criterio 15). También encontraron que cinco de los seis criterios adicionales que incluyeron en su experimento resultaron significativos (expresión de inseguridades, estilo de reporte, justificación de la falta de memoria y clichés). No obstante, uno de estos seis criterios (repeticiones) resultó significativo, pero en contra de las predicciones se presentó en mayor medida en las narraciones de sujetos sinceros (para una descripción detallada de los criterios adicionales incluidos en este estudio ver Köhnken et al., 1995). Así mismo, Vrij, Akehurst, Soukara y Bull (2004a) reportaron que los criterios 1, 3, 4, 5 y 6 fueron los más efectivos para diferenciar entre testimonios veraces y falsos.

La gran mayoría de los estudios que analizan la eficacia del CBCA ha empleado muestras de sujetos europeos o sujetos de raza blanca. Por este motivo, Ruby y Brigham (1998) investigaron las diferencias que podían surgir entre sujetos pertenecientes a distintos grupos étnicos. Los autores partieron de la idea de que existen diferencias a nivel verbal, en cuanto al estilo y contenido, entre las narraciones aportadas por individuos de diferentes razas. Plantearon como hipótesis que, debido a que el CBCA se desarrolló en una cultura europea de raza blanca, la aplicación de la prueba sólo sería eficaz para discriminar entre los testimonios realizados por personas de esta raza y que las narraciones realizadas por personas de raza negra incluirían significativamente menos criterios. Los resultados reportaron que el CBCA funcionaba diferente según la raza de la persona y que distintos criterios fueron mejores predictores de veracidad para un grupo étnico que para otro. En las narraciones veraces realizadas por sujetos de raza negra, los criterios que más se presentaron fueron el 3, 6, 12, 14, y 17

en comparación con las narraciones veraces de los sujetos de raza blanca. En las narraciones veraces proporcionadas por sujetos de raza blanca en comparación con las narraciones de sujetos de raza negra no existió ningún criterio que se presentara de forma significativa con mayor frecuencia. Considerando los testimonios de los sujetos de ambas razas, se observó que determinados criterios se presentaban significativamente más en las declaraciones veraces (criterios 2, 5, 7, 8, 9, 14 y 15). Sin embargo, en las narraciones falsas también hubo una mayor presencia de ciertos criterios: 1, 4, 11, 12 y 17. En cuanto a la clasificación de las declaraciones los resultados no fueron positivos. Encontraron que si se tomaban como veraces todas aquellas declaraciones en las que se encontraban presentes 5 criterios, el porcentaje de clasificación correcto para las narraciones veraces era del 89%, sin embargo, con esta regla un alto porcentaje de narraciones falsas eran clasificadas también como veraces (92%). Cuando emplearon criterios de decisión más rígidos (tomando en cuenta la presencia de 6 o 7 criterios), el número de narraciones veraces correctamente clasificadas descendía y el número de narraciones falsas correctamente clasificadas ascendía, es decir, con este criterio menos narraciones falsas eran clasificadas como verdaderas.

Por otro lado, los estudios también han revelado que existen diferencias en las puntuaciones del CBCA cuando se comparan los resultados de niños y adultos. Así lo revela un estudio realizado por Vrij et al. (2004a). También se ha encontrado que el conocimiento previo del contenido de los criterios del CBCA influye de forma negativa en la validez del instrumento y que los sujetos instruidos para mentir que anticipadamente conocen los criterios pueden proporcionar declaraciones que podrían aparentar ser verdaderas (Vrij, Akehurst, Soukara y Bull, 2002).

La eficacia del CBCA también ha sido comparada con otros procedimientos de evaluación del contenido de las declaraciones y se ha investigado si la combinación de estos métodos mejora la clasificación de los testimonios, concretamente es el *reality monitoring* (Jonson y Raye, 1981) el procedimiento con el que más se le ha comparado. El *reality monitoring*, que, procediendo de la investigación básica, tuvo más tempranas aplicaciones clínicas, ámbito en el que ha desarrollado una abundante investigación (véase por ejemplo, Bentall, Baker y Havers, 1991; Brebion, Smith, Gorman y Amador, 1997; Harvey, 1985; Johnson, Raye, Hasher y Chromiak,

1979; Raye y Johnson, 1980) y que posteriormente fue aplicado al ámbito forense (véase para una revisión Mitchell y Johnson, 2000), postula que los recuerdos de lo realmente visto tienen unas características distintas de lo no visto. Las autoras plantean que las memorias de origen externo tendrían más atributos contextuales y sensoriales, serían más detalladas semánticamente y tendrían menos información de operaciones cognitivas que las memorias de origen interno (Jonson y Raye, 1981).

Un estudio que contrasta los resultados del CBCA y del *reality monitoring* es el realizado por Sporer (1997). En su experimento empleó una muestra de 40 estudiantes de psicología (20 varones y 20 mujeres). A los sujetos se les daba la instrucción de narrar dos experiencias personales: una que fuera real y otra falsa. El objetivo del autor era comprobar la eficacia del CBCA y el *reality monitoring* en la discriminación de narraciones fabricadas y veraces y si el empleo de ambas pruebas mejoraba la clasificación de dichas narraciones. Los resultados demostraron que el CBCA fue efectivo en el 65% del total de las clasificaciones, con un 70% de eficacia en la clasificación de las narraciones verdaderas y un 60% de eficacia en la clasificación de las narraciones falsas. En lo que concierne al *reality monitoring*, el 71,3% de las declaraciones fueron correctamente clasificadas. De las declaraciones verdaderas el 75% fueron correctamente clasificadas y de las narraciones falsas el 67,5% fueron correctamente clasificadas. Al integrar el CBCA y el *reality monitoring* el porcentaje de clasificación incrementó a un 79%. En un estudio posterior, Vrij, Akehurst, Soukara y Bull (2004b) encontraron que los testimonios veraces obtenían mayores puntuaciones tanto en el CBCA como en los criterios del *reality monitoring*, con una eficacia de clasificación del 60% y el 74% respectivamente. No obstante, al integrar los resultados de ambas pruebas no encontraron mejoras y el porcentaje de clasificación se mantuvo en un 74%, en esta ocasión no se encontraron mejoras al integrar ambas técnicas.

Varios estudios también han planteado como opción combinar los procedimientos de evaluación del contenido verbal de las declaraciones con indicadores conductuales del engaño. Vrij et al. (2001) encontraron que las personas que mentían obtenían menores puntuaciones en los criterios del CBCA y en lo del *reality monitoring* y presentaban en mayor medida determinadas conductas que eran indicativas de engaño, como esperar mucho tiempo antes de dar una respuesta, hablar más rápido,

etc. Incluso encontraron que tanto el CBCA como el *reality monitoring* eran las herramientas más sensibles en la detección del engaño en relación a otras variables como la conducta no verbal. En un estudio anterior, Vrij et al. (2000) encontraron que utilizando de forma conjunta indicadores no verbales e indicadores verbales del engaño (CBCA y *reality monitoring*) el porcentaje de clasificaciones correctas incrementaba. Estos resultados fueron ratificados en un estudio más reciente por Vrij et al. (2004a) al examinar la conducta verbal y no verbal de niños y adultos.

### CONCLUSIONES

El Análisis de Contenido Basado en Criterios (CBCA) aún dista mucho de ser una herramienta completamente eficaz en la detección de testimonios engañosos y aún queda mucho por refinarla. Independientemente de que esta técnica se aplique a niños o a adultos existen muchos factores que influyen de forma negativa y que pueden modificar sus resultados. Como muestran las investigaciones, existen diferencias individuales, como la edad, la habilidad verbal, la actitud del entrevistador (Santtila et al., 2000), la familiaridad del evento (Blandon-Gitlin et al., 2005; Pezdek et al., 2004), el conocimiento previo de la prueba, (Vrij et al., 2002) y el grupo étnico de la persona (Ruby y Brigham, 1998), que hay que tener en cuenta y controlar en la medida de lo posible cuando se emplee esta técnica y que, por consiguiente impiden la inmediata aplicación individual del CBCA.

Por otro lado, y aunque los estudios demuestran que las declaraciones veraces contienen un mayor número de criterios frente a las declaraciones fabricadas, la principal y mayor desventaja del CBCA es que no existe un consenso general que establezca un número mínimo de criterios que debe incluir una declaración para ser catalogada como creíble y el peso que cada uno de ellos debe recibir. Landry y Brigham (1992) han propuesto como mínimo la presencia de cinco criterios para que una declaración sea catalogada como veraz. Sin embargo otros autores han manipulado en su experimento el número de criterios en la clasificación de las declaraciones y sus resultados no fueron tan positivos como esperaban (Ruby y Brigham, 1998). Así mismo, aún queda por definir el número de criterios que deben incluirse en la valoración del testimonio de adultos. Debido a que el CBCA se desarrolló para valorar las declaraciones de menores, es probable que algunos criterios no funcionen con adultos, como es el caso

del criterio 10 (incomprensión de detalles relatados con precisión). En esta misma línea, también hace falta estudios encaminados a definir un grupo de criterios aplicables a los testimonios de adultos. En cuanto al peso que debe recibir cada criterio, las investigaciones distan aún más en cuanto al establecimiento de un parámetro general. Sin embargo, se ha encontrado que ciertos criterios discriminan mejor que otros entre testimonios veraces y falsos (Ruby y Brigham, 1997, 1998). Posiblemente algunos de estos criterios deban recibir una mayor puntuación en la valoración general del CBCA, no obstante es indispensable realizar futuras investigaciones que nos ayuden a esclarecer esta problemática.

De acuerdo a lo expuesto, una alternativa viable en la detección de testimonios engañosos es la combinación de varias técnicas, como lo son el *reality monitoring* y los indicadores conductuales del engaño (Vrij et al., 2001; Vrij et al., 2000) que, como expusimos, en la mayoría de las ocasiones mejoran la clasificación de las declaraciones.

Con base en los inconvenientes antes mencionados, lo que sí queda claro es que el CBCA debe considerarse exclusivamente como un instrumento de apoyo y nunca como herramienta única sobre la cual se base la toma de decisiones judiciales (Santtila et al, 2000), o al menos no por el momento.

### REFERENCIAS

- Alonso-Quecuty, M.L. (1999). Evaluación de la credibilidad de las declaraciones de menores víctimas de delitos contra la libertad sexual. *Papeles del Psicólogo*, 73, 36-40.
- Bentall, R.P., Baker, G.A. y Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology*, 30, 213-222.
- Blandon-Gitlin, I., Pezdek, K., Rogers, M. y Brodie, L. (2005). Detecting deception in children: an experimental study of the effect of event familiarity on CBCA ratings. *Law and Human Behaviour*, 29, 187-197.
- Brebion, G., Smith, M.J., Gorman, J.M. y Amador, X. (1997). Discrimination accuracy and decision biases in different types of reality monitoring in schizophrenia. *Journal of Nervous and Mental Disease*, 185, 247-253.
- Harvey, P.D. (1985). Reality monitoring in mania and schizophrenia: the association of thought disorder and performance. *Journal of Nervous and Mental Disease*, 173, 67-73.

- Horowitz, S.W. (1991). Empirical support for statement validity assessment. *Behavioural Assessment*, 13, 293-313.
- Jonson, M.K. y Raye, C.L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Johnson, M.K., Raye, C.L., Hasher, L. y Chromiak, W. (1979). Are there developmental differences in reality monitoring?. *Journal of Experimental Child Psychology*, 27, 120-128.
- Köhnken, G., Schimossek, E., Aschermann, E. y Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, 80, 671-684.
- Lamb, M.E., Sternberg, K.J., Esplin, P.W., Hershkowitz, I., Orbach, Y. y Hovav, M. (1997). Criterion-based content analysis: a field validation study. *Child, Abuse & Neglect*, 21, 255-264.
- Landry, K.L. y Brigham, J.C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behaviour*, 16, 663-676.
- Masip, J., Sporer, S.L., Garrido, E. y Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law*, 11, 99-122.
- Mitchell, K. y Johnson, M. (2000). Source monitoring. Attributing mental experiences. En E. Tulving y F.I.M. Craik (Eds.), *The Oxford handbook of memory*. Nueva York: Oxford University Press.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G.S., Quas, J.A., Saywitz, K.J., Bidrose, S., Pipe, M.E., Rogers, M. y Brodie, L. (2004). Detecting deception in children: event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology*, 89, 119-126.
- Raskin, D.C. y Esplin, P.W. (1991). Statement Validity Assessment: interview procedures and content analysis of children's statements of sexual abuse. *Behavioural Assessment*, 13, 265-291.
- Raye, C.L. y Johnson, M.K. (1980). Reality monitoring vs. discriminating between external sources of memories. *Bulletin of the Psychonomic Society*, 15, 405-408.
- Ruby, C.L. y Brigham, J.C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: a critical review. *Psychology, Public Policy, and Law*, 3, 705-737.
- Ruby, C.L. y Brigham, J.C. (1998). Can criteria-based content analysis distinguish between true and false statements of african-american speakers?. *Law and Human Behaviour*, 22, 369-388.
- Santtila, P., Roppola, H., Runtti, M. y Niemi, P. (2000). Assessment of child witness statements using criteria-based content analysis (CBCA): the effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime & Law*, 6, 159-179.
- Sporer, S.L. (1997). The less travelled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373-397.
- Steller, M. (1989). Recent developments in statement analysis. En J.C. Yuille (Ed.), *Credibility assessment* (pp. 135-154). Dordrecht, Países Bajos: Kluwer.
- Steller, M. y Köhnken, G. (1989). Statement analysis: credibility assessment of children's testimonies in sexual abuse cases. En D.C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp.217-245). Nueva York: Springer.
- Undeutsch, U. (1989). The development of statement reality analysis. En J.C. Yuille (Ed.), *Credibility assessment* (pp. 101-121). Dordrecht, Países Bajos: Kluwer.
- Vrij, A. (2000). *Detecting lies and deceit: the psychology of lying and the implications for professional practice*. Chichester: John Wiley & Sons.
- Vrij, A., Akehurst, L., Soukara, S. y Bull, R. (2002). Will the truth come out? the effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behaviour*, 26, 261-283.
- Vrij, A., Akehurst, L., Soukara, S. y Bull, R. (2004a). Detecting deceit via analysis of verbal and nonverbal behaviour in children's and adults. *Human Communication Research*, 30, 8-41.
- Vrij, A., Akehurst, L., Soukara, S. y Bull, R. (2004b). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science*, 36, 113-126.
- Vrij, A., Edward, K. y Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin*, 27, 899-909.
- Vrij, A., Edward, K., Roberts, K. y Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behaviour. *Journal of Nonverbal Behaviour*, 24, 239-263.

# LA SIMULACIÓN DE ENFERMEDAD FÍSICA O TRASTORNO MENTAL

## MALINGERING OF ORGANIC ILLNESS OR MENTAL DISORDER

**Mercedes Inda Caro\***, **Serafín Lemos Giráldez\***, **Ana María López Rodrigo\***  
**y José Luis Alonso Rionda\*\***

\* Facultad de Psicología. Universidad de Oviedo. \*\* Servicios de Salud Mental de Asturias

Se analiza la naturaleza y la frecuencia de la simulación de síntomas clínicos, y se describen algunos instrumentos de medida generales, que han sido utilizados para determinar la validez de los síntomas, así como pruebas específicas para tomar decisiones respecto a la existencia de simulación en el ámbito clínico. En particular, se presentan algunos criterios de utilidad para evaluar la existencia de simulación de un trastorno de estrés post-traumático, de un síndrome orgánico cerebral, de amnesia y de un trastorno psicótico.

The purpose of this article is to analyze the nature and incidence of malingering, to describe several general measures used to determine symptom validity, as well as some instruments specifically created to assess deception or malingering in clinical settings. Particularly, some useful criteria in detecting the faking of post-traumatic stress disorder, organic brain syndromes, amnesia, and psychotic disorders, are discussed.

**E**l término "simulación", en su vertiente psicológica, está lleno de matices por cuanto alude a actitudes de encubrimiento (en el inglés británico, descritas como *dissimulation* o *deception*), de fingimiento o engaño (en el inglés americano, *faking*), o bien de invención consciente y deliberada de un trastorno mental o físico (en inglés, *malingering*), o de una incapacidad producida por un accidente o enfermedad, que en realidad no fueron causantes de esta, y de la que se deriva alguna ventaja personal. Esta es la acepción utilizada en el DSM-IV cuando la define como "la producción intencionada de síntomas físicos o psicológicos desproporcionados o falsos, motivados por incentivos externos como no realizar el servicio militar, evitar un trabajo, obtener una compensación económica, escapar de una condena criminal u obtener drogas. Bajo algunas circunstancias, la simulación puede representar un comportamiento adaptativo: por ejemplo, fingir una enfermedad mientras se está cautivo del enemigo en tiempo de guerra" (American Psychiatric Association, 1995, p. 698).

Este concepto es matizado por Resnick (1997) cuando diferencia entre simulación pura (*pure malingering*) o fingimiento de un trastorno no existente; simulación parcial (*partial malingering*) o la exageración consciente de síntomas presentes o de un trastorno ya superado; y falsa imputación (*false imputation*), la atribución errónea de síntomas reales a una determinada causa, debido a

un engaño inconsciente o a una mala interpretación de la situación.

Gorman (1982) diferenció también entre el acto y el estado de simular, por cuanto un acto implica voluntad, actitud asertiva de deseo y de propósito; mientras que el estado, desde un punto de vista legal, sería inherente a la persona, debido a su condición social o a posibles limitaciones.

Por otro lado, se plantea la cuestión de si la simulación puede ser, en sí misma, reflejo de algún trastorno mental. Ello puede resultar bastante claro en el llamado trastorno "facticio" o ficticio (American Psychiatric Association, 1994), en donde la persona finge síntomas físicos o psicológicos intencionadamente, con el fin de asumir el papel de enfermo; en el trastorno histriónico de la personalidad, por la falta de control sobre su conducta manipuladora; pero también pudiera ser un comportamiento neurótico la exageración consciente de molestias físicas o mentales para conseguir un objetivo económico, laboral, profesional, etc., ya que ninguna persona en sus cabales suele llegar a esos extremos, ni escogería rutas tan tortuosas y dolorosas, para obtener posibles ganancias. Sin embargo, desde el Derecho, no se hacen estas matizaciones y se considera que ciertos comportamientos tienen una intención clara por parte de la persona que los lleva a cabo (Gorman, 1982).

La incidencia de la simulación después de un accidente no es muy conocida, estimándose entre el 1 y el 50% de los casos (Henderson, 1986; Miller y Catledge, 1972), dependiendo de si la fuente de información es el abogado del demandante o las compañías de seguros. Un factor in-

Correspondencia: Mercedes Inda Caro. Facultad de Psicología. Universidad de Oviedo. Plaza Feijóo, s/n. 33003 Oviedo. España. E-mail: indamaria@uniovi.es



fluyente son las condiciones económicas y laborales, ya que se ha observado que aumenta cuando se esperan despidos inminentes en una empresa y disminuye cuando las personas mejoran su situación financiera o laboral. La disparidad respecto a la incidencia de la simulación puede deberse a la dificultad existente a la hora de diferenciar entre quienes inventan completamente los síntomas y las personas que exageran molestias ya existentes; en cuyo caso, podría hablarse más bien de "pacientes".

Por otra parte, la identificación de simuladores y la investigación sobre estos comportamientos han sido tradicionalmente muy complicadas. Así, por ej., en el ámbito de la reclamación de indemnizaciones por secuelas derivadas de algún accidente, al igual que en procesos judiciales en donde se reclaman derechos o la exención o atenuación de responsabilidades, se comprueba que las personas que tienen acceso a información relacionada con las alteraciones cerebrales y las pruebas neuropsicológicas pueden alterar con mayor probabilidad sus resultados (Coleman, Rapport, Millis, Ricker y Farchione, 1998; Youngjohn, Lees-Hayley y Binder, 1999). Los estudios de simulación han mostrado que el entrenamiento (como es proporcionar información de los déficit más comunes asociados con una lesión cerebral o llamar la atención sobre la presencia de medidas en la detección de posible simulación) es otro de los factores que puede afectar a la validez de los índices de simulación. Los estudios que se realizan con estudiantes universitarios, entrenados en simular una enfermedad orgánico-cerebral, muestran que sus ejecuciones se parecen más a las de los pacientes reales que la que realizan simuladores novatos o sin experiencia en el campo de la evaluación neuropsicológica. Sin embargo, su ejecución es exageradamente peor que la que realizan personas con verdaderas lesiones cerebrales.

Una de las muchas limitaciones observadas en este tipo de investigaciones es la excesiva utilización de muestras de estudiantes universitarios, a quienes se les invita a simular un daño cerebral (Strauss et al., 2002; Vickery, Berry, Inman, Harris y Orey, 2001), por cuanto no son comparables con los verdaderos simuladores, ya que estos buscan un beneficio económico por su "lesión", suelen tener un amplio conocimiento del problema por haber estado expuestos a múltiples evaluaciones por diferentes expertos, frecuentemente repetidas en intervalos de entre una y dos semanas, han observado a otros pacientes como ellos, pasando por un entrenamiento inadvertido e inconsciente, y suelen tener más tiempo para preparar su futura evaluación. Obviamente, los verdaderos simuladores no reconocen dicha condición y, en consecuencia, difícilmen-

te son incluidos en estudios de investigación. Pese a estas dificultades, en los estudios con análogos es posible abordar la simulación a través de las inconsistencias en las ejecuciones repetidas de una prueba. De hecho, es bastante difícil mantener la misma ejecución cuando se emplean baterías de gran amplitud, por lo que este método permite obtener un índice válido para detectar una posible simulación (Cullun, Heaton y Grant, 1991). Cuando se comparan pacientes implicados en litigios por sus lesiones con pacientes con las mismas lesiones pero sin demandas judiciales, se ha encontrado que en el primer grupo hay menor consistencia entre sus evaluaciones que en el segundo grupo, siendo los resultados de las últimas evaluaciones peores que los de las primeras evaluaciones (Reitan y Wofson, 1996, 1997).

## TRASTORNOS QUE SUELEN SER OBJETO DE SIMULACIÓN

### *Trastorno de estrés post-traumático (TEPT)*

Para establecer el TEPT es necesario hacer una descripción metódica de los síntomas, los tratamientos previamente aplicados, y una cuidadosa corroboración sobre la veracidad de la información. En la fase de obtención de información, el clínico debe ser muy cuidadoso de no proporcionar información alguna a la persona sobre cuales son los síntomas claves de este trastorno. Además, si el clínico comienza la evaluación cuestionando las respuestas del paciente, la naturaleza agresiva de la evaluación podría afectar al estilo de respuesta y a la posibilidad que la persona intentase justificar su daño con la presencia de síntomas extremos. Uno de los inconvenientes que tiene el diagnóstico clínico es que se basa en el auto-informe del paciente sobre los síntomas subjetivos; por lo que la actividad que tenía la persona supuestamente afectada una semana antes a la ocurrencia del estresor deberá ser comparada con la actividad que mantiene en el momento de la evaluación, y examinar si existe una razonable relación entre los síntomas y el estresor, el tiempo transcurrido entre el estresor y los síntomas desarrollados, y la relación entre algún trastorno previo y los síntomas actuales. El psicólogo debe insistir en que el sujeto proporcione una descripción detallada de los síntomas del trastorno. Los simuladores pueden que tengan un amplio conocimiento acerca de cuales son los síntomas característicos que configuran el TEPT, pero normalmente fallan en adecuar esos síntomas a su vida cotidiana dando una descripción poco detallada. Los síntomas inventados suelen ser vagos o bastantes artificiosos y forzados (Pitman, Sparr, Saunders y McFarlane, 1996). Otra indicación que habrá que seguir

para considerar una posible simulación es si la persona minimiza otras posibles causas de sus síntomas y exagera como causa de sus síntomas el accidente o situación por la que solicita una compensación.

Resnick (1997) sugiere que terceras personas no deberían estar presentes durante la evaluación, por dos razones: la primera, porque la presencia de parientes o amigos cercanos pueden ser empleados como fuentes "independientes" para corroborar la veracidad de los síntomas; y la segunda porque es más fácil para el clínico hacer frente a un posible simulador cuando se enfrenta a él solo. Otro aspecto que señala este autor es la conveniencia de que el clínico adopte una postura amable y cordial a la hora de comunicar al sujeto que se sospecha que puede estar exagerando los síntomas, y no hacerlo con una actitud que exprese agresividad o humillación, empleando, p. ej., testigos; ya que de este modo puede negarse a admitirlo y mostrar enfado. Resnick establece algunos criterios a seguir por el clínico que sospecha de la existencia de un TEPT simulado (Tabla 1), y un modelo de decisión clínica para determinar la existencia de simulación de este trastorno (Tabla 2).

### **Síndrome de daño cerebral post-traumático**

Este trastorno es bastante frecuente en la sociedad actual, en gran medida como consecuencia de accidentes laborales y de tráfico. Se suele manifestar con dolores de cabeza, mareo, estados de ansiedad, inestabilidad emocional, visión borrosa, déficit de concentración y problemas de memoria. De todos los síntomas, los más fáciles de simular son los emocionales. El daño cerebral post-traumático puede ser confundido con el TEPT, al ser también bastante frecuente después de sufrir un traumatismo craneoencefálico. Ambos trastornos tienen componentes comunes como son la amnesia de algún elemento del hecho traumático, síntomas depresivos (anhedonia, afecto restringido, actitud pesimista hacia el futuro), alteraciones del sueño, irritabilidad, dificultades de concentración e intolerancia a los sonidos fuertes; sin embargo, algunos autores, como Price (1994), sostienen que no es posible que ambos trastornos se encuentren en la misma persona porque, quien ha sufrido una lesión cerebral con pérdida de la conciencia, no va a poder reexperimentar el acontecimiento traumático, de modo que el carácter excluyente de ambos trastornos justificaría la conclusión de simulación cuando se presentan conjuntamente.

### **Amnesia**

Las principales medidas desarrolladas para detectar la simulación de trastornos de la memoria incluyen test muy

simples, que pueden ser realizados correctamente incluso por personas con daño cerebral, en los que los simuladores suelen mostrar mayores déficits que los verdaderos pacientes. Se denomina *efecto suelo* cuando el simulador novato exagera su papel y comete muchos fallos en estas pruebas; sin embargo, Cercy, Schretlen y Brandt (1997) señalan varios problemas en el empleo de estas técnicas. En primer lugar, las personas con experiencia en simular síntomas amnésicos se dan cuenta de esta estrategia y evitan ejecutar demasiado mal las pruebas. En segundo lugar, a pesar de la aparente simplicidad de las pruebas, algunos pacientes con verdaderos daños cerebrales o con trastornos neuro-psiquiátricos tienen bastante dificultad en ejecutarlas correctamente.

**TABLA 1**  
**CRITERIOS A SEGUIR EN LA EVALUACIÓN DE UN POSIBLE CASO DE SIMULACIÓN DEL TEPT (RESNICK, 1997). LA PRESENCIA SIMULTÁNEA DE DOS O MÁS DE ESTAS CARACTERÍSTICAS SUGIERE QUE LA PERSONA SE ENCUENTRA EN EL UMBRAL DE UNA SIMULACIÓN**

1. Registro de un pobre rendimiento.
2. Anteriores situaciones de "incapacitación".
3. Discrepancia entre el rendimiento mostrado y el realizado en su tiempo libre.
4. Sueños repetitivos e invariables.
5. Rasgos de personalidad antisocial.
6. Funcionamiento excesivamente bueno antes del accidente o suceso traumático.
7. Actitud evasiva ante ciertas preguntas.
8. Descripción inconsistente de los síntomas.

**TABLA 2**  
**MODELO DE DECISIÓN CLÍNICA PARA ESTABLECER UN DIAGNÓSTICO DE SIMULACIÓN DEL TEPT (RESNICK, 1997)**

- A. Establecimiento de un motivo lógico y plausible para simular un TEPT.**
- B. Presencia de, al menos, dos de los siguientes criterios:**
1. Situación laboral irregular o insatisfacción en el trabajo.
  2. Reclamaciones anteriores por daños.
  3. Capacidad para realizar actividades en su tiempo libre, pero no en el trabajo.
  4. Ausencia de pesadillas, o presencia de pesadillas que son exactas al hecho traumático sucedido.
  5. Rasgos de personalidad antisocial.
  6. Actitud evasiva y contradicciones.
  7. Actitud de no cooperación durante la evaluación.
- C. Confirmación de simulación por una de estas dos situaciones:**
1. Admisión de estar simulando los síntomas.
  2. Evidencia psicométrica inequívoca de simulación o fuerte corroboración de estar simulando los síntomas.

Una nueva corriente se está desarrollando para poder detectar la simulación de los trastornos amnésicos, representada por el análisis del fenómeno de la interferencia proactiva. Este fenómeno sucede cuando una información aprendida interfiere en la adquisición o el recuerdo posterior de un nuevo mensaje, y ello se analiza, por ejemplo, mediante el paradigma de aprendizaje de una lista de palabras, según su posición serial en la misma. La interferencia proactiva se manifiesta como un declive en el recuerdo, a medida que se va avanzando en la lista de palabras, de modo que las primeras palabras aprendidas interfieren la codificación y almacenamiento de las últimas. La interferencia es mayor cuando la nueva información es muy similar a la ya almacenada anteriormente, como sucede con palabras que pertenecan a la misma categoría semántica, que cuando se presenta una categoría de palabras diferentes de las previamente almacenadas; lo que produce cierta recuperación en los procesos de memoria (Wikens, 1970). El efecto de la interferencia proactiva se ha constatado en pacientes con verdaderas lesiones cerebrales, pero no en las personas que quieren simular daños en su memoria, recordando mejor las últimas palabras de la lista que las primeras. Este fenómeno, sin embargo, no ha sido confirmado por otros autores, al no haber encontrado diferencias de grado entre simuladores y pacientes con un genuino daño cerebral (Baker, Hanley, Jackson, Kimance y Slade, 1993). El interés por utilizar la interferencia proactiva como detector de simulación parte del supuesto de que es este un proceso cognitivo automático, que está fuera del control consciente del sujeto.

Baker y cols. (1993) investigaron también la influencia que tendría un distractor en el recuerdo de un conjunto de estímulos cuando aparece entre la exposición estos y la tarea de recuerdo; pero no encontraron diferencias significativas cuando el recuerdo de los ítem se solicitaba después de un intervalo de 20 segundos en los que lo sujetos debían contar hacia atrás. Los falsos pacientes, en cambio, ejecutaron mucho peor esta prueba.

Otros estudios se han centrado en la detección de la simulación a través de pruebas de memoria implícita (García Domingo, Gregredo López y Fernández Guinea, 2004). Las personas amnésicas generalmente muestran una cercanía a las personas normales en la ejecución de test con el efecto del *priming*, y en tareas que no requieren un recuerdo explícito del episodio aprendido. Por ejemplo, cuando un paciente con amnesia procesa una serie de palabras sin anunciarle que posteriormente se le va a pedir que las recuerde, y más adelante se le presenta la raíz de la palabra o fragmentos de la misma, la

probabilidad que recuerden la palabra correcta es muy grande. Este fenómeno de *priming* se considera que está controlado por procesos neurocognitivos independientes. Wiggins y Brandt (1988) sospecharon que los simuladores realizarían más pobremente estas pruebas de memoria implícita que los verdaderos pacientes, y, efectivamente, encontraron que la ejecución de los simuladores fue comparativamente peor, pero no encontraron diferencias estadísticamente significativas.

Otro procedimiento para detectar a los simuladores ha sido valorando lo que se ha llamado *feeling-of-knowing*, o la sensación del individuo acerca de si tiene un recuerdo parcial; es decir, si es consciente que sufre amnesia; sin embargo, algunos autores consideran limitado este índice, debido a la variabilidad existente entre los verdaderos pacientes amnésicos. Sí se ha encontrado que personas que simulaban trastornos de memoria y que habían ejecutado peor un test de elección forzada mostraron niveles bajos en el *feeling-of-knowing* (Schacter, Harbluck y McLachlan, 1984).

### **Psicosis**

La prevalencia del fingimiento de la psicosis es desconocida, pero Resnick (1984) considera que, debido al movimiento de la desinstitucionalización, podría ir en aumento, ya que miles de enfermos crónicos, que preferirían vivir en un ambiente más protegido, se encuentran en la actualidad en situación de marginalidad social. Con los drásticos recortes en los programas sociales y las mejoras en las condiciones de hospitalización, las personas con trastornos mentales podrían exagerar sus síntomas para conseguir una ayuda médica; como es el caso de los pacientes con esquizofrenia, que muestran notable habilidad para parecer sanos o enfermos según los objetivos que tengan en el momento (Rogers, Kropp, Bagby y Dickens, 1992). La simulación de un trastorno psicótico puede responder a diversas razones: evitar responsabilidades en el caso de estar implicado en juicios penales; tiempo atrás, evitar realizar el servicio militar o librarse de un destino peligroso; obtener algún beneficio económico por secuelas o daños psicológicos; o dejar de permanecer en prisión, simulando un estado psicótico para conseguir el traslado a un hospital, para obtener más fácilmente drogas o para tener más oportunidades de escapar.

Los especialistas se lamentan de la falta de criterios diagnósticos para establecer la existencia de simulación en estos casos. Resnick (1997), no obstante, sugiere algunos principios a tener en cuenta por el clínico que sospecha de un caso de simulación. Así, respecto a las

alucinaciones auditivas, los simuladores deberán ser preguntados acerca de las estrategias que emplean para reducir o hacer desaparecer las voces. Además de que los verdaderos pacientes suelen tener una disminución de este tipo de alucinaciones cuando la esquizofrenia se encuentra en una fase de remisión mientras que en los brotes agudos se dan con mucha frecuencia, las estrategias de afrontamiento que emplean los pacientes con esquizofrenia incluyen realizar actividades específicas como trabajar o ver la televisión, cambiar la postura, por ejemplo tumbarse o andar, mantener una conversación con una persona cercana, o tomar rápidamente la medicación. En general, comprueban que sus alucinaciones suelen disminuir cuando se implican en cualquier actividad. Estas acciones espontáneas y el correspondiente efecto amortiguador de las alucinaciones, debería evaluarse en los supuestos simuladores; ya que si la persona no tiene un conocimiento profundo de la enfermedad, durante la entrevista no proporcionará esta información. Las alucinaciones genuinas se caracterizan por un amplio rango de susurros a gritos sin sentido; sin embargo la cadencia del discurso es normal. Por el contrario, los simuladores a veces refieren el contenido de sus alucinaciones con un lenguaje artificioso y demasiado complicado. Respecto a las alucinaciones visuales, señala también Resnick que cuando el contenido es dramático o atípico debe ser sospechoso de simulación.

En relación con los delirios, los simuladores refieren su aparición de repente, cuando es sabido que un delirio real se va construyendo a lo largo de meses o años, hasta su sistematización. Cuando aparecen las ideas delirantes, estas suelen tener poca influencia en la vida diaria del paciente, aunque esté convencido de la veracidad de las mismas. En la valoración del carácter genuino de un delirio, Resnick indica que hay que tener en cuenta su contenido. El contenido de los delirios fingidos suele ser persecutorio, y en ocasiones de grandiosidad, pero raras veces auto-despreciativos. Además, el comportamiento de los simuladores no suele ser acorde con el supuesto delirio, mientras que en las personas con una verdadera psicosis es mayor la relevancia conductual.

Otro síntoma que suelen presentar las personas con un trastorno psicótico es la conducta de mutismo. El mutismo puede aparecer como síntoma aislado o como parte de una psicosis simulada. La conducta catatónica o la flexibilidad cética son muy difíciles de mantener durante un periodo prolongado; de modo que una manera de comprobar si una persona está fingiendo puede ser observar cómo reacciona a un pinchazo en la espalda. Las personas con un verdadera catatonía responderán de la

misma manera independientemente si previamente ven o no el estímulo doloroso; mientras que los simuladores responderán de manera diferente dependiendo si anticipan o no el estímulo doloroso. Si la persona ve al examinador aproximarse con un alfiler, mostrará una pequeña reacción, ya que previamente habrán tensado los músculos; pero si se le pincha sin haber visto el alfiler mostrará menor contracción muscular y dilatación pupilar.

En el caso de los trastornos de conversión, es más difícil determinar si se está produciendo una situación de simulación. Resnick señala que el criterio diferencial es si la conducta de mutismo se encuentra bajo el control voluntario de la persona. Conocer los detalles exactos a cerca de cómo la persona dejó de hablar es muy importante, a juicio de este autor. Las personas con un trastorno de conversión suelen ser capaces de escribir y de susurrar, y suelen tener historia de otros síntomas de conversión, como un trastorno disociativo; mientras que es más frecuente hallar en la historia del simulador conducta antisocial con frecuentes mentiras o un historial criminal.

Respecto a la simulación de una depresión psicótica, es sabido que la variación diurna de los síntomas forma parte de su manifestación clínica; mostrando el paciente genuino mayor gravedad de los síntomas y estados de ánimo más disfóricos por las mañanas y una cierta mejoría al final de la jornada. Esta oscilación clínica es más probable que no sea referida por los simuladores, al carecer de un conocimiento profundo del trastorno.

## MÉTODOS DE EVALUACIÓN

Los procedimientos de valoración de la simulación, en el ámbito clínico, se han basado en el uso de medidas neuropsicológicas convencionales y en el empleo de test específicos para tal fin.

Dentro de la primera opción, se han analizado las curvas de ejecución en tareas de dificultad variable (Baker et al., 1993; Frederick, Crosby y Wynkoop, 2000; Tehula y Sweet, 1996), los aciertos en pruebas de recuerdo, de reconocimiento y tareas en las que se exige discriminar entre dos tipos de estímulos (Coleman et al., 1998; Slick, Iverson y Green, 2000; Suhr y Gunstad, 2000; Sweet et al., 2000), tareas de memoria (Davis, King, Bajzar y Squire, 1995; Hanley, Baker y Ledson, 1999), la amplitud de memoria para los dígitos (Strauss et al., 1999; Suhr, Tranel, Wefel y Barrash, 1997); la comparación de índices de atención e índices de memoria (Mittenberg, Azrin, Millsaps y Heilbronner, 1993), y el conocimiento semántico (Mittenberg, Theroux-Fichera, Heilbronner y Zielinski, 1995). Si bien pruebas de este tipo se han considerado muy óptimas para detectar posi-

bles casos de simulación, se contempla la necesidad de aplicar pruebas complementarias para mejorar la validez y fiabilidad de los resultados. Lezak (1995) refiere los siguientes test neuropsicológicos clásicos para la detección de posibles simuladores:

- Test de Bender, con la recomendación de llevar a cabo un retest, transcurridos bastantes días desde la primera evaluación, ya que al sujeto se le olvidará cuáles eran los patrones de respuestas, e invirtiendo el orden de las tarjetas.
- Test de Retención Visual Benton, en donde los simuladores cometen más errores de distorsión que los pacientes con lesiones cerebrales, pero no más errores de omisión.
- La Batería de Halstead-Reitan (incluyendo el WAIS), en la que los simuladores ejecutan peor los test que los pacientes lesionados, a excepción del Test de Categorías, el Test de Ejecución Tactil y la parte B del Test de Trazado (*Trail Making Test*).
- El MMPI, en donde también los simuladores obtienen peores perfiles que los verdaderos pacientes.
- El PICA (*Porch Index of Communicative Ability*), para la simulación del trastorno afásico.

Entre las pruebas diseñadas específicamente para evaluar la simulación, se advierten dos vertientes metodológicas. Una se basa en el llamado “paradigma de validación de síntomas” (Pankratz, Fausti y Peed, 1975), originalmente diseñado para evaluar los déficit en el funcionamiento sensorial, siendo luego extendido para detectar el fingimiento de lesiones relacionadas con la memoria (Binder y Willis, 1991; Frederick y Foster, 1991; Iverson, Franzen y McCracken, 1991; Pankratz, 1983). Este paradigma implica la administración de pruebas de elección forzada de dos alternativas de respuesta, que basan los resultados en términos de probabilidades (Slick, Hopp, Strauss y Thompsom, 1997; Tombaugh, 1996), y en las que se establecen unos intervalos de confianza, por debajo o por encima de los cuales se consideran las puntuaciones como indicativas de fingimiento o exageración de síntomas y se establecen puntos de corte a partir de los cuales se seleccionan las respuestas. Por ejemplo, en una persona que no está intentando simular, sus respuestas deberán tener al menos un 50% de aciertos, que es el resultado esperado cuando se responde al azar. Este fue el primer sistema de corte empleado, pero se encontró que, cuando se ensayaba con personas normales que fingían ser simuladores, estas no puntuaban por debajo de los niveles de respuesta esperados por azar, aunque sí tenían mayor número de errores que los verdaderos pacientes dañados y honestos; por ello, se optó por establecer pun-

tos de corte en relación con la ejecución que se espera de una persona con una verdadera lesión y sin propósitos de exageración o fingimiento.

Algunos investigadores han empezado a explorar la utilidad de medidas encubiertas, obtenidas a partir de las respuestas “objetivas” que emiten los simuladores, que no puedan ser manipuladas por estas personas o “mejoradas” con las sucesivas evaluaciones. Un ejemplo de ello es la versión computerizada del *Portland Digit Recognition Test* de Rose, Hall y Szalda-Petree (1995), en la cual incluye una medida de la latencia de respuesta del sujeto. Estos autores encontraron que la incorporación de esta medida a la versión original creada por Binder y Willis (1991) mejoraba la sensibilidad de la prueba en la identificación de los posibles simuladores.

La segunda vertiente metodológica consiste en el estudio del tipo de respuesta que ejecuta el paciente; por ejemplo, la manera en que el paciente lee palabras muy sencillas o cuenta el número de puntos que aparecen en una pantalla (Boone et al., 2000; Strauss et al., 2002). Un ejemplo sería el *Dot Counting Test* (Binks, Gouvier y Waters, 1997), prueba en la que al sujeto se le presenta una serie de tarjetas con puntos agrupados y sin agrupar y se pide que cuente el número de puntos que ve en la pantalla, se mide el número de aciertos y el tiempo empleado en contar los estímulos.

### EVALUACIÓN MEDIANTE ENTREVISTAS Y AUTOINFORMES

Otra manera de detectar la simulación consiste en valorar la sintomatología conductual del problema. Las primeras aproximaciones se hicieron mediante pruebas cuyo objetivo específico no era la detección del engaño, pero que incluían alguna subescala para medir la validez del instrumento. El primero de ellos y el más conocido es el *Minnesota Multiphasic Personality Inventory* (MMPI), cuya Escala F tiene el fin de detectar estilos atípicos de respuesta. Con esta escala, sin embargo, se detectaron varios problemas, como el solapamiento entre las puntuaciones obtenidas por verdaderos pacientes y las obtenidas por los posibles simuladores, y la falta de sensibilidad de esta escala para detectar situaciones concretas de simulación como, por ejemplo, la simulación de trastornos de memoria. También se ha puesto en cuestión la utilidad de la subescala DM del 16 PF de Cattell para estos fines.

El *M Test* (Beaber, Marson, Michelli y Millis, 1985) fue el primer instrumento desarrollado con el objetivo primario de detectar posibles simuladores. Sin embargo, los estudios de validación también han puesto en entredicho



esta prueba, indicando Hankins, Barnard y Robbins (1993) que parece detectar más bien a personas que manifiestan déficit o deterioro cognitivo. En un intento de mejorar este test, Rogers y cols. desarrollaron un nuevo sistema de puntuación para este test obteniendo resultados óptimos en la diferenciación entre pacientes psiquiátricos de un correccional y de un hospital (Rogers, Bagby y Gillis, 1992). Smith, Forum y Schinka (1993), sin embargo, no confirmaron estos resultados con una población similar.

La *Malingering Scale* (Schretlen, 1986) ha sido otro intento de construir una prueba para detectar a los simuladores mediante una evaluación psicométrica. Este instrumento consta de dos escalas: la escala de simulación de deficiencia mental (*malingering retardation, MgR*), y la escala de simulación de locura (*malingering insanity, Mgl*). Sin embargo Smith y Burger (1997) indican que los estudios que se han desarrollado para la validación de la prueba tienen déficit metodológicos, relativos a la selección sesgada de la muestra, así como prácticos, como la longitud de la prueba (150 ítem) y la necesidad de contar con un evaluador experimentado para la aplicación de la misma.

Rogers y cols. han desarrollado la SIRS (*Structured Interview of Reported Symptoms*) con el objetivo de desenmascarar a las personas que están fingiendo o exagerando algún trastorno mental (Rogers, Gillis y Bagby, 1990; Rogers, Gillis, Dickens y Bagby, 1991; Rogers, Kropp et al., 1992). La SIRS es una entrevista que consta de 172 preguntas, distribuidas en 8 escalas primarias y 5 escalas adicionales o complementarias. Las primeras se emplean para diferenciar entre personas que son honestas en sus respuestas y simuladores, así como en la descripción de los estilos de respuestas. Las escalas primarias exploran: Síntomas raros (*Rare Symptoms, RS*) (genuinos, pero poco comunes); Síntomas improbables y absurdos (*Improbable and Absurd Symptoms, IA*); Combinación de Síntomas (*Symptoms Combination, SC*) (en referencia a la poca probabilidad de que dos síntomas verdaderos se produzcan simultáneamente); Síntomas Obvios (*Blatant Symptoms, BL*); Síntomas Sutiles (*Subtle Symptoms, SU*) (en referencia a síntomas que son vistos por los pacientes como problemas, cuando en realidad no lo son); Síntomas selectivos (*Selective Symptoms, SEL*); Síntomas graves (*Severity Symptoms, SEV*); y Síntomas referidos frente a observados (*Reported versus Observed Symptoms, RO*).

Con las escalas complementarias el clínico puede realizar una interpretación de los estilos de respuesta del sujeto. Este segundo bloque lo forman: Valoración directa de

honestidad (*Direct Appraisal of Honesty, DA*); Actitud de estar a la defensiva (*Defensive Symptoms, DS*); Síntomas iniciales (*Symptom Onset, SO*); Síntomas demasiado específicos (*Overly Specified Symptoms, OS*); Síntomas de inconsistencia (*Inconsistency of Symptoms, INC*).

Las preguntas se pueden agrupar en tres categorías: (a) preguntas detalladas, que se dirigen a explorar la gravedad de sintomatología específica; (b) preguntas repetidas, de control, sobre las respuestas a las preguntas del apartado a; y (c) preguntas generales, que tratan de investigar patrones de síntomas y problemas psicológicos.

Para la creación de la SIRS, Rogers (1984) realizó una revisión literaria para identificar estrategias de posible utilidad para la detección de simuladores. A partir de aquí, seleccionó las estrategias que cumplían los siguientes criterios: (a) las que parecían relevantes para detectar la simulación de una enfermedad mental en contraposición a otras formas de engaño, y (b) aquellas estrategias que eran fáciles de estandarizar. Basándose en estos criterios, Rogers seleccionó quince estrategias, a partir de las cuales generó 330 preguntas, que conformaron la primera versión de la SIRS. La construcción formal de las escalas se llevó siguiendo dos pasos: por un lado, mediante el consenso de ocho expertos en simulación se buscó la validez aparente y validez descriptiva de las escalas propuestas, asignando los ítems a la estrategia que creían más adecuada, de modo que cuando al menos cinco de los ocho expertos coincidían con la clasificación de Rogers el ítem era situado en la escala correspondiente; y, por otro lado, se calculó las correlaciones ítem-escala, eliminándose aquellos ítems que no correlacionaban con la escala asignada. Los coeficientes alpha de las escalas se situaron entre 0.66 y 0.92, con una media de 0.86 (Rogers, 1997).

Los resultados para cada una de las escalas son clasificados en cuatro categorías: honestos, indeterminados, probable fingimiento o engaño definitivo. Se considera que la persona está intentando engañar si la puntuación en tres o más de las escalas primarias se sitúa en el rango de fingimiento probable; o si la puntuación total del SIRS (la suma de los ítems de las preguntas generales y las detalladas) excede de 76. Se considerará que está siendo honesta si la puntuación en seis o más de las escalas primarias está en el rango de *honesto* o la puntuación global es igual o inferior a 71.

Finalmente, la SIMS (*Structured Inventory of Malingered Symptomatology*) (Smith y Burger, 1997) es otro instrumento para evaluar la simulación, que consiste en un autoinforme de 75 ítem dicotómicos (verdadero/falso), agrupados en cinco escalas desarrolladas para detectar

posible falsación en las cinco condiciones clínicas más comunes de simulación: baja inteligencia, trastornos afectivos, daño neurológico, psicosis y amnesia. Se obtiene una puntuación total a partir de las cinco escalas. Los ítem fueron obtenidos de dos fuentes diferentes: la primera, a partir de instrumentos ya existentes como el MMPI, el SIRS y el WAIS-R, los cuales han demostrado cierta utilidad en la detección de una posible simulación. Estos ítem fueron modificados para aumentar su sensibilidad en la detección de situaciones concretas de simulación. Y una segunda fuente fue a partir de las características cualitativas de personas simuladoras (Resnick, 1984; Rogers, 1984; Seamons, Howell, Carlisle y Roe, 1981).

### CONCLUSIONES

La simulación, el engaño o el fingimiento, potencialmente se pueden manifestar en todo tipo de enfermedades somáticas y trastornos mentales. Por eso, es necesario utilizar también procedimientos de evaluación diferentes para desenmascarar a las personas que pudieran presentar o exagerar síntomas de muy diverso orden; ya que no es lo mismo simular daño físico, como una lesión cerebral, que daño psicológico, como un trastorno mental. Cualquiera que sea la naturaleza de los síntomas, con frecuencia es conveniente contar con la participación del psicólogo, a la hora de evaluar su autenticidad; y en particular cuando las molestias referidas por la persona afectan a las funciones cognitivas, como la atención o la memoria, y son susceptibles de evaluación neuropsicológica. A pesar de que las exploraciones médicas pueden descartar daños o alteraciones orgánico-cerebrales, en ocasiones el paciente no deja de referir problemas en su vida diaria a la hora de conducir, de acordarse de cosas, etc., siendo estos los argumentos empleados por los peritos en las diferentes causas judiciales y sobre los que es necesario tomar una decisión.

La credibilidad sobre el carácter verdadero o simulado de los síntomas o de estos testimonios, en cuanto que entra dentro de las opiniones subjetivas, difícilmente puede ser objeto de investigación científica. Sin embargo, sí puede ser analizada la validez de los síntomas o del cuadro clínico que la persona manifiesta, para determinar con criterios científicos una probable situación de fingimiento o exageración. En este sentido, se sugiere hacer una aproximación multifactorial para determinar la existencia de una situación de simulación; para lo cual, es necesario: (1) determinar la gravedad del daño, mediante un registro de los diferentes síntomas; (2) evaluar al paciente mediante pruebas estandarizadas; (3)

considerar diagnósticos psicológicos o médicos alternativos para explicar la causa de los síntomas aducidos por la persona; (4) utilizar pruebas que se ajusten a las características demográficas de los sujetos evaluados; y (5) utilizar, simultáneamente, pruebas neuropsicológicas e indicadores de validez específicos, para determinar la posible simulación de los síntomas.

Esbec Rodríguez y Gómez-Jarabo (1999) han descrito, por ejemplo, hasta veinte características que pueden indicar la existencia de simulación de un trastorno mental; de entre las cuales, destacan dos como las más importantes: la presencia de algún beneficio o ganancia externa clara por el padecimiento de estos síntomas y la constatación que el sujeto ya había presentado previamente síntomas similares a los alegados en el momento actual.

La Psicología Forense, en consecuencia, tiene en España un importante reto por delante, en lo que se refiere a la determinación, con base científica, de la validez de los testimonios y de la sintomatología aducida por personas que están en proceso de enjuiciamiento o que han sido víctimas de accidentes o agresiones. En particular, es necesario desarrollar procedimientos estructurados y estandarizados, que permitan emitir juicios fundamentados sobre la posible existencia de simulación. Los procedimientos desarrollados por Arce y Fariña (2005), en este sentido, son un ejemplo a seguir en la consecución de este objetivo.

### REFERENCIAS

- American Psychiatric Association. (1995). *Manual Diagnóstico y Estadístico de los Trastornos Mentales* (4<sup>ed.</sup>). Barcelona: Masson.
- Arce, R. y Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG). *Papeles del Psicólogo*, 26, 59-77
- Baker, G. A., Hanley, J. R., Jackson, H. F., Kimrance, S. y Slade, P. (1993). Detecting faking of amnesia: Performance differences between simulators and patients with memory impairment. *Journal of Clinical and Experimental Neuropsychology*, 15, 668-684.
- Beaber, R. J., Marson, A., Michelli, J. y Millis, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. *American Journal of Psychiatry*, 142, 1478-1481.
- Binder, L. M. y Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, 3, 175-181.
- Binks, P. G., Gouvier, W. D. y Waters, W. F. (1997).

- Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology*, 12, 41-46.
- Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., et al. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b test. *Archives of Clinical Neuropsychology*, 15, 227-242.
- Cercy, S. P., Schretlen, D. y Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomena. En R. J. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 85-107). New York: Guilford Press.
- Coleman, R. D., Rapport, L. J., Millis, S. J., Ricker, J. H. y Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, 20, 201-210.
- Cullun, C., Heaton, R. y Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H. O. Doerr y A. S. Carlingf (Eds.), *Forensic neuropsychology: Legal and scientific bases* (pp. 141-171). New York: Guilford Press.
- Davis, H. P., King, J. K., Bajszar, J. H. y Squire, L. R. (1995). *Colorado Malingering Test Package, version 2.0*. Colorado Springs, CO: Colorado Neuropsychology Tests.
- Esbec Rodríguez, E. y Gómez Jarabo, G. (1999). Signos de sospecha en la simulación de los trastornos mentales. *Revista Española de Psiquiatría Forense, Psicología Forense y Criminología*, 8, 35-44.
- Frederick, R. L., Crosby, R. D. y Wynkoop, T. E. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology*, 15, 281-300.
- Frederick, R. L. y Foster, H. G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment*, 3, 596-602.
- García Domingo, L., Gregredo López, L. y Fernández Guinea, S. (2004). Evaluación de la simulación de problemas de memoria dentro del ámbito legal y forense. *Revista de Neurología*, 38, 766-774.
- Gorman, W. F. (1982). Defining malingering. *Journal of Forensic Sciences*, 27, 401-407.
- Hankins, G. C., Barnard, G. W. y Robbins, M. L. (1993). The validity of the M Test in a residential forensic facility. *The Bulletin of the American Academy of Psychiatry and Law*, 21, 111-121.
- Hanley, J. R., Baker, G. A. y Ledson, S. (1999). Detecting the faking of amnesia: A comparison of the effectiveness of three different techniques for distinguishing simulators from patients with amnesia. *Journal of Clinical and Experimental Neuropsychology*, 21, 59-69.
- Henderson, J. (1986). *Psychic trauma claims in civil and administrative law*. Paper presented at the 17th American Academy of Psychiatry and the Law Meeting, Philadelphia, PA.
- Iverson, G. L., Franzen, M. D. y McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior*, 16, 667-676.
- Lezak, M. D. (1995). *Neuropsychological assessment (3rd ed.)*. New York: Oxford University Press.
- Miller, H. y Catledge, N. (1972). Simulation and malingering after injuries to the brain and spinal cord. *Lancet*, 1, 580-585.
- Mittenberg, W., Azrin, R., Millsaps, C. y Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. *Psychological Assessment*, 5, 34-50.
- Mittenberg, W., Theroux-Fichera, S., Heilbronner, R. y Zielinski, R. E. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology*, 11, 271-281.
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills*, 57, 367-372.
- Pankratz, L., Fausti, A. y Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology*, 43, 412-422.
- Pitman, R. K., Sparr, L. F., Saunders, L. S. y McFarlane, A. (1996). Legal issues in posttraumatic stress disorder. En B. A. van der Kolk, A. C. McFarlane y L. Weisaeth (Eds.), *Traumatic Stress* (pp. 378-397). New York: Guilford Press.
- Price, K. P. (1994). Posttraumatic stress disorder and concussion: A they incompatible? *Defense Law Journal*, 43, 113-120.
- Reitan, R. M. y Wofson, D. (1996). The question of validity of neuropsychology test scores among head-injured litigants: Development of a Dissimulation Index. *Archives of General Neuropsychology*, 2, 573-580.
- Reitan, R. M. y Wofson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: Development of the Retest Consistency Index. *Clinical Psychologist*, 11, 69-76.
- Resnick, P. J. (1984). The detection of malingered mental

- illness. *Behavioral Sciences and the Law*, 2, 20-38.
- Resnick, P. J. (1997). Malingering of posttraumatic disorders. In R. J. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 130-152). New York: Guilford Press.
- Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law*, 2, 93-112.
- Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Rogers, R., Bagby, R. M. y Gillis, J. R. (1992). Improvements in the M Test as a screening measure for malingering. *The Bulletin of the American Academy of Psychiatry and Law*, 20, 101-104.
- Rogers, R., Gillis, J. R. y Bagby, R. M. (1990). Cross validation of the SIRS with a correctional sample. *Behavioral Sciences and the Law*, 8, 85-92.
- Rogers, R., Gillis, J. R., Dickens, S. E. y Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the SIRS. *Psychological Assessment*, 3, 89-96.
- Rogers, R., Kropp, P. R., Bagby, R. M. y Dickens, S. E. (1992). Faking specific disorders: A study of the Structured Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology*, 48, 643-648.
- Rose, F. E., Hall, S. y Szalda-Petree, A. D. (1995). Computerized Portland Digit Recognition Test: The measurement of response latency improves the detection of malingering. *Clinical Neuropsychologist*, 9, 124-134.
- Schacter, D. L., Harbluck, J. L. y McLachlan, D. R. (1984). Retrieval without recognition: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior*, 23, 593-611.
- Schretlen, D. (1986). *Malingering: Use of a psychological test battery to detect two kinds of simulation*. Ann Arbor, MI: University Microfilms International.
- Seamons, D. T., Howell, R. J., Carlisle, A. L. y Roe, A. V. (1981). Rorschach simulation of mental illness and normality by psychotic and nonpsychotic legal offenders. *Journal of Personality Assessment*, 45, 130-135.
- Slick, D. J., Hopp, G., Strauss, E. y Thompsom, G. (1997). Victoria Symptoms Validity Test. Odessa, FL: Psychological Assessment Resources.
- Slick, D. J., Iverson, G. L. y Green, P. (2000). California Verbal Learning Test indicators of suboptimal performance in a sample of head-injury litigants. *Journal of Clinical and Experimental Neuropsychology*, 22, 509-579.
- Smith, G. P., Borum, R. y Schinka, J. A. (1993). Rule-out and rule-in scales for the M Test for malingering: A cross-validation. *The Bulletin of the American Academy of Psychiatry and Law*, 21, 107-110.
- Smith, G. P. y Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *The Bulletin of the American Academy of Psychiatry and the Law*, 25, 183-189.
- Strauss, E., Hultsch, D. F., Hunter, M., Slick, D. J., Patry, B. y Levy-Bencheton, J. (1999). Using intraindividual variability to detect malingering in cognitive performance. *Clinical Neuropsychologist*, 14, 420-432.
- Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. y Hultsch, D. F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology*, 17, 423-444.
- Suhr, J. A. y Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, 15, 415-424.
- Suhr, J. A., Tranel, D., Wefel, J. y Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, 19, 500-514.
- Sweet, J. J., Wolfe, P., Sattleberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., et al. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, 15, 105-114.
- Tehula, W. N. y Sweet, J. J. (1996). Double cross-validation of the Vooklet Category Test in detecting malingered traumatic brain injury. *Clinical Neuropsychologist*, 10, 104-116.
- Tombaugh, T. N. (1996). *Test of memory malingering*. New York: MHS.
- Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J. y Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, 16, 45-73.
- Wiggins, E. C. y Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behavior*, 12, 57-78.
- Wikens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77, 1-15.
- Youngjohn, J. R., Lees-Hayley, P. R. y Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology*, 14, 511-515.



# EL ENGAÑO Y LA MENTIRA EN LOS TRASTORNOS PSICOLÓGICOS Y SUS TRATAMIENTOS

## DECEPTION AND LYING IN PSYCHOLOGICAL DISORDERS AND THEIR TREATMENTS

**Manuel Porcel Medina\* y Rubén González Fernández\*\***

\*Departamento de Orientación Laboral de la Unión General de Trabajadores. \*\*Centro Privado de Atención Psicológica

*Este artículo pretende examinar los fenómenos de la mentira, el engaño y el autoengaño como términos que están directamente relacionados con los problemas psicológicos y sus tratamientos. Veremos que, en numerosas ocasiones, el engaño y el autoengaño no son más que la cobertura psicológica o ficticia de los problemas de la vida. Pero como los problemas psicológicos sólo pueden neutralizarse a través de psicoterapia, veremos que toda psicoterapia que se precie deberá incorporar en sus juegos terapéuticos a la mentira. Y el terapeuta ejercitará esta mentira de la psicoterapia, de una manera muy efectiva, a través de la técnica que conocemos como "intención paradójica". En efecto, será esta técnica intrincada en la mentira la que pueda contrarrestar la propia mentira del cliente aquejado de un problema psicológico. La intención paradójica funcionará cuando se administre acompañada con el movimiento del cliente, moldeando dicho movimiento en el camino que el propio cliente estará transitando en cada momento.*

*This article tries to examine lie, faking, and self-deception phenomena as directly related terms to psychological problems and its treatments. We will realize that, in many occasions, faking and self-deception are not more than a psychological or fictitious coverage of life problems. Inasmuch as psychological problems can only be neutralized by means of psychotherapy, we will see that all self-respecting psychotherapies ought to incorporate the lie in its therapeutic games. And the therapist will put into practice this lie of the psychotherapy, in a very theatrical way, by means of the technique known as "paradoxical intention". Indeed, this lie-involved technique will be the best way to counteract self-deception in a psychologically disordered client. Paradoxical intention will work when paired up with client's movement, shaping this movement through the way the client concerned will cover at a given moment.*

**E**xisten diversas frases que, de forma condensada y efectista, pretenden definir al hombre: "el hombre es un animal que usa herramientas"; "el hombre es un animal que posee la capacidad de lenguaje"; "el hombre es un animal que come pan". Nuestra contribución al catálogo de frases es la siguiente: "el hombre es un animal que miente". Es evidente que otros animales utilizan las apariencias para sobrevivir, pero es quizás el hombre el único animal que trata con la mentira de manera reflexiva, es decir, el único capaz de usar la mentira en relación a su persona, a su identidad. Si un camaleón camufla su cuerpo los humanos serían capaces de camuflar hasta lo más profundo de sus entrañas. La sugestión es un prodigio de la mentalidad genuinamente humana. Sólo el hombre puede ver fantasmas.

Aunque excede en mucho los austeros objetivos de este artículo, queremos manifestar que la importancia de la mentira, en cuanto que se puede relacionar con la su-

gestión, es máxima, porque detrás se esconde el fundamento de la propia psicología: el yo, la reflexividad.

### LA VERDAD DE LA VIDA Y TEXTURA VITAL DE LOS PROBLEMAS PSICOLÓGICOS

Vamos a discutir este punto de la exposición ayudándonos de una metáfora y esto lo vamos a hacer así por dos motivos fundamentales. El primero es por un motivo didáctico y facilitador de la comprensión y de la explicación. El segundo es en honor a la psicoterapia que nosotros ejercitamos, que no es otra que una actualización o acomodación de una psicoterapia genérica conocida como Terapia de Aceptación y Compromiso. Y si utilizamos una metáfora, le hacemos honor a esta terapia porque uno de sus principios generales es el convencimiento de que las metáforas son verdaderas herramientas terapéuticas en el contexto clínico. Se utilizan para hacer ver al cliente una serie de conceptos y de fenómenos que, de otra manera, serían mucho más difícil de discernir.

La metáfora que vamos a utilizar es la del "camino". Y no vamos a pensar en un camino poético, poco transita-

Correspondencia: Rubén González Fernández.  
E-mail: rgonzalez@correo.cop.es



do y perdido en un paraje que, precisamente, por estar perdido, tiene un encanto y misterio particulares. No va a ser de esta especie. El camino que vamos a usar como metáfora es un camino moderno, urbano, con su rotondas o glorietas, con sus centros comerciales a los lados, con cruces de caminos muy transitados y regulados por semáforos, pasos de peatones, cedas el paso, etc.

Y tiene que ser un camino moderno, urbano, porque los problemas psicológicos han surgido en un contexto moderno y urbano que no es otro que el contexto de la ciudad. En las sociedades arcaicas había poco espacio para los problemas psicológicos en tanto que la vida estaba muy normalizada. Era una sociedad cerrada y esto significa que el modo de vida y los problemas derivados de tal modo de vida estaban estrictamente pautados. La ciudad, en cambio, es en su génesis un cruce de modos de vida. Podríamos decir que ya no es la familia la que regula los modos de vida sino el "mercado". Ahora todo se comercializa: el alimento, la ropa, los enseres de la vida. Pero también se mercadea con los modos de vida y se ofrecen diferentes alternativas profesionales, familiares, de ocio, que no sólo son alternativos sino que además, en muchos casos, son incompatibles y contradictorios. Y el problema de tener diferentes modos de vida disponibles para usar es que el individuo empieza a ejercitar la "responsabilidad de elegir". Podríamos decir, y ya adelantamos la que será nuestra tesis más fuerte, que los problemas psicológicos van a tener que ver, de una manera más o menos directa u oblicua, con esta responsabilidad de elegir.

Volviendo a nuestra metáfora del camino moderno y urbano, imaginemos que la *vida* es una rotonda. Una rotonda tiene la función de repartir el tráfico. Es la ordenación de un cruce de caminos y es la parte del camino urbano que nos facilita dejar una dirección para coger otra.

¿Qué ocurriría si no existiesen estas rotondas? Probablemente uno dejaría un camino y emprendería otro sin más dilación, de manera directa, sin una estación de tránsito. ¿No habéis dado alguna vez más de una vuelta a nuestra rotonda para aclarar las ideas en cuanto a la dirección que tenemos que tomar? Es una especie de tiempo extra-decisional. Pero, ¿y si permanecemos en la rotonda por tiempo indefinido sin tomar ninguna ruta? Estos son para nosotros *los problemas psicológicos*. Es un "atasco vital" en la decisión de qué camino emprender a partir de un determinado momento. Tenemos varios alternativos y todos ellos presentan sus ventajas y sus inconvenientes. Unos son muy fáciles de recorrer pero nos llevan a lugares poco atractivos. Otros son muy

duros, con mucho tráfico, con semáforos, no obstante nos llevan a lugares más atractivos que los anteriores. Pero no tomamos una decisión, estamos "atascados". Mientras sólo uno de nosotros esté atascado en la rotonda, no hay mayor problema. El problema crece cuando la rotonda se colapsa y ya no permite que otros la utilicen adecuadamente. Es cuando el problema psicológico individual se convierte en un problema de calado social. Y el problema también aumenta cuando, al estar permanentemente en la rotonda, no llegamos a ningún destino, no cumplimos con nuestro trabajo en la ciudad, no cumplimos con nuestras obligaciones familiares, etc.

A tal fin se ha creado todo un entramado circulatorio alternativo y ajeno a la vida que nos permite seguir conduciendo, aunque a ningún destino. Conducimos en círculo porque necesitamos movernos. No tenemos más objetivo que permanecer en pausa el tiempo que necesitamos para volver a nuestro camino urbano ordinario.

Este entramado circulatorio alternativo, ajeno a la vida y estéril es la *psicologización*. Y no nos referimos solo a la psicologización que se ejercita mediante el verbo, sino también aquella que se practica con fármacos, con flores –léanse las "flores de Bach" por ejemplo-, con magia, con futurología. Todo este entramado está al servicio de aquellos que quedaron atrapados en la rotonda indefinidamente y que no lograron tomar una decisión vital.

La psicoterapia se ha convertido en una institución intermedia (Pérez Álvarez, 1999), entre instituciones que fallan. Cuando la gente se colapsa entre varias alternativas de vida y ninguna de ellas le satisface, surge una vía encubierta, con sentido momentáneo pero sin sentido último que son los problemas psicológicos. La institución psicoterapéutica ofrece la cobertura necesaria a los problemas psicológicos dando una explicación funcional. Una persona puede vivir entre las alternativas conflictivas de querer estar delgada y atracarse de comida. El colapso vital de esta persona estaría en diferir infinitamente la decisión que ha de tomar por uno u otro camino. Atiborrarse de comida no importando las consecuencias o intentar llevar una alimentación adecuada. El camino intermedio (la rotonda de la vida psicopatológica) suele ser atiborrarse y purgar lo ingerido. Soportar la certeza de que tomar uno u otro camino depende de uno mismo es, a todas luces, insostenible. Es así como se alimenta la tercera vía muerta y vacía de llamar a esta ausencia de decisión personal: "problema psicológico". Empieza aquí todo un entramado normativo que da cobijo y explicación al comportamiento tan irracional de atiborrarse y purgar lo ingerido. El problema psicológico se llama bulimia y el profesional que ha

de acabar con ella es el psicólogo o psiquiatra. Justo desde este momento la única persona que podría dar un golpe de efecto a esta situación –la persona con bulimia– queda subyugada a la institución psicoterapéutica que ya hace las cosas por ella. Es precisamente esta pérdida de responsabilidad lo que hace que los problemas de la vida, los atolladeros coyunturales, se conviertan en problemas psicológicos, en callejones sin salida estructurales.

Pero claro, por paradójico que parezca, es únicamente desde la psicoterapia desde donde se puede restablecer la cordura perdida. A saber, es a través de psicoterapia como hemos de revertir los problemas psicológicos en problemas de la vida, en pruebas o dificultades que requieren una continua toma de decisiones por parte del que vive. Cuando la toma de decisión se enquistaba –y nos atrapa la rotonda– estamos alimentando un problema hasta hacerlo psicológico. Para decirlo con una frase, hemos de despsicologizar al cliente desde la propia psicología. Despsicologizar significa, en este contexto, eliminar la cobertura psicológica, por tanto externa e inabordable por parte del cliente, de los problemas de la vida. Las decisiones las tiene que tomar el cliente y el deber del psicólogo es precisamente no caer en la tentación de hacerle llegar las suyas propias. Utilizando la conjugación que permite una explicación cabal y comprensiva de los problemas psicológicos, “autor-actor” (Quiroga Romero, 1999), decimos que la psicoterapia desde los quicios que manejamos aquí (contextual, conductual, ontológica) es sencillamente el intento de hacer pasar al cliente por el gradiente desde una irresponsabilidad e indecisión de la vida que está viviendo siendo meramente actor de un papel, hasta la responsabilidad y toma de contacto real con la vida que le ha tocado vivir llegando a ser el propio autor de la misma. Cómo lograr este cambio es algo que intentamos explicar más abajo.

### EL AUTOENGAÑO COMO MODULADOR DE LOS TRASTORNOS PSICOLÓGICOS

Un aspecto relevante en la casuística psicopatológica se nos revela con el fenómeno del autoengaño. Si el engaño en general es un elemento de interacción social absolutamente extendido en el funcionar de las sociedades modernas, que va desde los aspectos económicos y políticos hasta los más íntimos, el autoengaño, se podría decir, resulta ser la extensión de-generada que acaba por impregnar de caminos obtusos la conciencia y la voluntad personal.

Podríamos definir el autoengaño como un modo de dirigir nuestra vida cuando no sólo ignoramos lo que com-

porta la dirección elegida sino, sobre todo, cuando ignoramos o pretendemos ignorar que en realidad, sin remedio, hemos tomado algún camino que tiene aparejado una serie de consecuencias. Este es el sentido que ya le daba Plutarco al término, cuando expresaba que el autoengaño era algo más que la incapacidad de reconocer que no sabemos de muchas cosas, pues al fin lo dramático es que no sabemos lo que somos. Si la incertidumbre y la inseguridad nos paraliza y por prudencia decidimos estancarnos, cabe preguntarse si la parálisis no será también una opción acompañada de riesgos y, por tanto, en ocasiones, carente de prudencia.

Una mentira puede tener distintas variantes. Puede ser inocente, humorística, puede ser algo perversa y puede ser hasta bondadosa y útil. El autoengaño, en cambio se podría decir que, sin perjuicio de su consideración como inocente, humorístico, perverso, piadoso y útil, no toma ninguna variante especial siendo acaso un poco de todo. El engaño apunta a un objetivo consciente, pero el autoengaño es inconsciente no sabiendo uno lo que hace, porque, como dijera el famoso escritor, “más de un espejo es un velo” (Wilde 1889). Los trastornos psicológicos exhiben esta peculiaridad en la mayoría de las ocasiones. Así, una persona afectada de anorexia ignora muchas veces su temor al rechazo público afanándose en adelgazar o peleándose con su cuerpo. Un neurótico con conductas compulsivas ignora que, tras la conducta de lavarse las manos constantemente y su pavor a la contaminación, se esconde su obstinación de no aceptar el carácter obligadamente incierto de la vida.

El problema principal que se encuentra un psicólogo suele ser el de mostrar claramente cuál es el verdadero problema, que por lo general está oculto a los ojos de quien lo padece.

De manera general, la persona con problemas psicológicos sufre por algo que ella misma excluye o aparta y con lo que, sin embargo, está comprometida. La voluntad consciente, diríamos, está entregada a la actitud de pelearse reflexivamente con los elementos psicológicos molestos y, por tanto, la distracción favorece el descuido en reconocer, y si acaso vencer, el verdadero problema que está afectando su vida. Como bien ha detectado la filosofía de la Terapia de Aceptación y Compromiso y otras filosofías y autores antes, lo que subyace a la dedicación o la lucha de una persona en el control de los síntomas emocionales y cognitivos (que paradójicamente acaba por alimentar) es una evitación existencial o vital (experiencial), una dificultad en aceptar las cosas que no se pueden cambiar (Luciano y Hayes, 2001).

Aparece aquí el fenómeno del autoengaño como algo crucial, en el sentido de que el esfuerzo en concentrarse en los elementos psicológicos acaba por ocultar los elementos sustanciales de un conflicto personal no resuelto (Fuentes Ortega, 1994) y eso termina por dar naturaleza psíquica a un problema que sólo el afrontamiento personal finalmente puede resolver. Un individuo que se queja de depresión puede ensombrecer así su responsabilidad en el afrontamiento del dolor, el sufrimiento y la pena por su insistencia de permanecer acostado, apático y poco activo. Pero sólo cuando se abandona a una continua tortura auto-inflingida, reprochándose su actitud deprimida, cristaliza el autoengaño o un verdadero problema psicológico (se cierra el círculo), pues en la reflexión crítica consigo mismo acaba por crear un espacio inerte, donde se dedica a intentar despejar un entramado psíquico cuyo último sentido es oscurecer el afrontamiento del verdadero problema. En este esquema el auto engañado acaba por perder ya la perspectiva del problema original y muchas veces reclama ayuda para salir de un circuito fatal que fue emprendido con la intención de calmar la percepción molesta de un conflicto, pero que acabará, posteriormente, por dejar a la persona sin capacidad de respuesta o ciego ante ese conflicto que, no por no percibir, deja de inquietar y en el fondo sostener y consolidar el malestar psicológico. Al respecto resulta sugerente el ejemplo que hace Paul Watzlawick cuando recuerda cómo la antropóloga Margaret Mead distinguía a los americanos de los rusos diciendo que mientras los primeros simulaban dolor de cabeza para eludir responsabilidades, los segundos, necesitaban realmente sufrir el dolor de cabeza para el mismo fin (Watzlawick, 1975). Pues bien, tal vez un problema psicológico sea, sobre todo, un dolor de cabeza "ruso" generado por uno mismo para pasar de puntillas por aspectos importantes de la vida, y ya sobrevenido, se agudiza cuando uno se dedica a intentar buscar analgésicos para un mal que la cefalea sólo trataba de sortear.

Al fin el autoengaño, como pretendemos dibujarlo, es perfectamente solidario con la idea del síntoma según ha sido descrito en un ensayo reciente (Pérez Álvarez, 2003), y comparte el trance de la expresión o manifestación de un problema real, pero también cumple la función de ser un intento adaptativo, una tregua o incluso una forma de vida.

### LA EXTRAÑA VERDAD DE LOS TRATAMIENTOS PSICOLÓGICOS

En multitud de ocasiones, un psicólogo tiene que disimular, callar, decir verdades a medias, hacer comentarios

sesgados que tienen que ver con el engaño, con la mentira o, cuando menos, con no mostrar la verdad desnuda. Unas veces por no herir la sensibilidad de su paciente y otras por mera educación. Sin embargo pensamos que la función simulativa y la apariencia juegan un papel sustancial y no sólo anecdótico en el trabajo terapéutico.

Utilizando un ejemplo de la medicina, diríamos que el cirujano puede operar sin la conciencia del paciente. La medicación, por lo demás, funciona con relativa independencia de las creencias reales de la persona que las toma, pero no ocurre lo mismo en el quehacer psicológico. Si algo caracteriza a la terapia psicológica es la importancia radical del fenómeno de la apariencia, hasta el punto de ser imposible desarrollarla sin aparentar, sin aparentarse el psicólogo al paciente y viceversa (incluso cabría dudar si una verdadera terapia se podría realizar sin conciencia de que tal proceso -el terapéutico- está en marcha). El médico puede estar ausente, el psicólogo tiene que estar, al menos, co-presente.

Bajo nuestro punto de vista, un tratamiento psicológico es algo similar al juego en un partido de fútbol (otras analogías similares han sido ensayadas anteriormente: la terapia como una partida de ajedrez y en general como un juego y como un desafío lleno de giros inesperados aparecen, por ejemplo, en las novelas del psicólogo existencialista Irvin D. Yalom (Yalom, 1992; 1996). El juego determinará la victoria aunque, para ganar, el juego debe desarrollarse en el marco de un partido que obliga a ciertas reglas, pero que nunca fija de antemano el éxito. El tratamiento psicológico se da en el marco de una ceremonia (García Sierra, 2001), la ceremonia psicoterapéutica, la podemos llamar. Respecto a la importancia del concepto de ceremonia para la psicología, se puede recurrir a la lectura de un trabajo significativo de Juan Fuentes y Ernesto Quiroga (Fuentes y Quiroga, 1998). Al decir esto, pretendemos simplemente indicar que la terapia siempre se desarrolla en el marco de una serie de secuencias transitorias que obedecen a ciertas normas: las sesiones tienen unos tiempos aproximadamente pautados tanto para el inicio como para su finalización, el psicólogo y el paciente pueden hablar en turnos sucesivos donde las correcciones se hacen más en un sentido que en otro, donde la autoridad le pertenece más a uno que a otro, etc.

Dicho esto, nuestra postura es la siguiente: un tratamiento o terapia es, sobre todo, una táctica puesta en juego, como lo es en el fútbol la táctica del entrenador, pero que debe ajustarse de continuo a las condiciones

reales de lo que ocurre en el campo, de lo que se habla en la consulta. Un psicólogo no puede aplicar sin más una serie de pasos hasta alcanzar un fin porque tiene que acompañar continuamente sus pasos con los del paciente, como lo haría un delantero con respecto al defensa del equipo contrario, y esto significa, digámoslo sin tapujos, que la terapia es un juego de mentiras arriesgadas donde el psicólogo tiene que mantener al paciente en el convencimiento de que le va a dar una solución a su problema, solución que decíamos antes, nunca tiene lograda previamente, que sólo será verdad en la medida que se consiga mantener durante el proceso la mentira de que se dispone de tal solución. Se puede resumir todo esto eufemísticamente: El psicólogo, si aspira a ser merecedor de tal título, debe mantener la credibilidad. Pero sin eufemismos añadimos: a través de distintos movimientos de prestidigitación.

El moldeamiento es un procedimiento utilizado por los terapeutas de conducta que consiste en partir de una serie de comportamientos previos e ir extendiendo progresivamente logros parciales hasta alcanzar un punto o logro final. Un agorafóbico puede, por ejemplo, aceptar circular por ciertos lugares pero no aceptaría otros en absoluto. El secreto, se podría decir, consiste en lograr que el paciente recorra los lugares que no aceptaría recorrer empezando por lograr que recorra aquéllos que acepta recorrer sin mucha resistencia. Si uno lo piensa detenidamente, se da cuenta de que lo que realmente está en juego es la voluntad del paciente y que para controlarla, de alguna manera, hay que someterla a engaño, porque realmente no está claro que la habituación no pudiera alcanzarse haciendo directamente lo que no se quiere hacer, pero puede resultar más efectivo conseguir que haga lo que no le importa hacer para, finalmente, indicarle hacer lo que no haría por mucho que se lo pidiésemos y que realmente es lo necesario para tener éxito. Las técnicas de aproximaciones sucesivas tienen este sentido. No es que un fóbico a los ascensores sea incapaz de subir a un sexto piso, es que el psicólogo tiene que conseguir que se decida a subir a dicho piso y, además, debe ser el propio paciente quien decida subir por el ascensor voluntariamente pese a acudir a la ayuda del profesional porque no está dispuesto a subir por voluntad propia. No es difícil adivinar que todo lo que hay por medio contiene mucho de creencia más que de realidad, pues el cliente tiene que atribuir al procedimiento un valor que no es estrictamente cierto: no es la habituación la que hace disminuir el miedo y por tanto le permite subir a un piso alto, es su decisión de subir la

que hace que el proceso de habituación se ponga en marcha y es con la libre decisión del paciente con la que no puede dejar de trabajar nunca un psicólogo.

Las técnicas *racionalistas* cognitivas también tienen su punto de pericia engañosa, pues se basan muchas veces en contrarrestar un rígido pensamiento catastrofista con evidencias contrapuestas igualmente sesgadas, porque lo cierto es que el discurso del terapeuta no es, muchas veces, más que un decir, que en realidad él puede readaptar a cada caso hasta el punto de decir una cosa y la contraria, según convenga, con el objetivo funcional de vencer una rigidez más que de convencer al pensante de otra verdad, lo que podría ser contraproducente.

Algunas técnicas como el *intercambio de papeles* dejan patente en qué consiste el juego terapéutico, a saber: encontrar la verdad a través de la apariencias. El psicólogo adopta el papel de paciente para que éste se de cuenta de que algún elemento de su discurso es un obstáculo para avanzar. Pero, ¿por qué no decirselo directamente? El sentido de la técnica es que el cliente se de cuenta sin sentirse agraviado o atacado de modo que reaccione defensivamente e ignore lo fundamental.

Todas las terapias, se podría decir, contienen elementos paradójicos y tanto en terapias conductistas, cognitivas y otras, abundan componentes de este orden. Pero a decir verdad, los apuntes anteriores con respecto al moldeamiento, los sesgos y el cambio de papeles habrían de interpretarse como casos particulares de una clase o modo general posible de abordar la terapia. En la literatura psicológica aparece incorporada, de hecho, una técnica aplicada con carácter plenamente consecuente, global y, creemos, particularmente exitosa si cumple la función para la que está diseñada y pensada; a saber, la técnica de la *intención paradójica*. Ya desde Adler, pasando por Víctor Frankl hasta la actual Terapia de Aceptación y Compromiso existe una tradición que ha hecho de ella todo un modo de proceder. En esta tradición psicológica la técnica mencionada puede interpretarse no ya tanto como una técnica más o como un aspecto residual de programa terapéutico, sino como toda una manera de tratar con los problemas psicológicos. Esa es, por otra parte, la postura comprometida en el presente trabajo.

La intención paradójica nos revela nitidamente el fenómeno del autoengaño presente en los trastornos psicológicos. Alguien con una disfunción eréctil puede desear las relaciones sexuales pero no quiere arriesgarse a quedar mal parado. Un fóbico social puede desear la relación con otras personas pero le cuesta aceptar la

posibilidad de tener contratiempos en las relaciones sociales. Alguien puede desear insinuarse a otra persona pero teme ponerse colorada. Hay quien desea adelgazar pero no le agrada la posibilidad de hacer ejercicio ni dieta. En todos estos casos la empresa del paciente acaba por enfocarse hacia la lucha contra los elementos secundarios de orden psicológico (miedo, angustia, rumiaciones...) no afrontando el conflicto original. De esta manera, la intención paradójica resuelve, o trata de resolver, el enredo con un disfraz, a saber, intentando involucrar al paciente en elementos secundarios paradójicos para así arrojarle al enfrentamiento con el conflicto base. Si producto de la angustia alguien comienza a rumiar el modo de no atascarse al hablar se pide al paciente que desee voluntariamente tropezar al hablar, ya que así el elemento rumiativo pierde sentido funcional, enfrentando a la persona con el conflicto de hablar aun a riesgo de tropezarse, lo que muy fácilmente ayudará a la fluidez del hablante.

La maniobra de la intención paradójica se basa en desacreditar al conflicto secundario (psicológico) intentando que el cliente se involucre en provocar un mal que él atribuye a una emoción o sentimiento y no a la voluntad de no afrontar. Así las cosas, si el paciente pone su voluntad a trabajar para sufrir los efectos psicológicos incómodos, estos desaparecerán.

La intención paradójica aprovecha el autoengaño de la persona que vive su problema (emoción como impedimento de la acción de afrontar) para alentar un logro psicológico (alivio cognitivo-emocional) a través de la acción. El éxito, como se entenderá, es más que probable, pues aunque el afrontamiento se ponga en marcha con intenciones de alivio psicológico, la acción está deshaciendo de facto el conflicto de base que explica todo el entramado. Lo que destruye la intención paradójica es la coartada de anteponer el contenido psicológico a la acción y esto lo consigue haciendo creer al paciente que mediante una acción paradójica, que por lo demás se mostrará efectiva, desaparecerá el contenido psíquico adverso. Y desaparecerá probablemente, pero, como decimos, porque deja de tener sentido utilitario para la persona al estar de hecho enfrentando el conflicto primario, es decir, corriendo el riesgo de tomar una dirección y no quedarse indefinidamente dando vueltas en la rotonda. Tampoco es asunto menor que la intención paradójica, por su sustancialidad e importancia al hacer figurar el conflicto de fondo, viene bien presentarla (esconderla) bajo el formato del humor consistente en reírse de uno mismo (Frankl, 1946)

Acabamos ya diciendo que lo escrito supone sólo un intento franco de reconocer que la mentira, la estrategia, la vía oblicua, son aspectos ineludibles y definidores de gran parte de lo que llamamos Psicología, lo que no hemos de ver necesariamente como despectivo. Acaso ya lo expresara de otro modo uno de los padres de nuestra disciplina, Alfred Binet, cuando para medir la inteligencia descubrió que lo que realmente hacía era asignar valor a los fallos y no a los aciertos: "mientras la Lógica se interesa por los procesos intelectuales que tienen que ver con la verdad, la Psicología se interesa especialmente por los procesos intelectuales que tienen que ver con el error".

## REFERENCIAS

- Frankl, V. (1946/2004). *El hombre en busca de sentido*. Barcelona: Editorial Herder.
- Fuentes Ortega, J.B. (1994) Introducción del "conflicto de normas irresuelto personalmente" como figura antropológica (específica) del campo psicológico. *Psicohema*, 6 (3), 421-446.
- Fuentes Ortega, J.B. y Quiroga Romero, E. (1998) A propósito de Mesmer y la Hipnosis. La distinción entre ceremonias abiertas y ceremonias cerradas. *Revista de Historia de la Psicología*, 19 (2/3), 421-429.
- García Sierra, P. (Ed.). (2001). *Diccionario filosófico "manual del materialismo filosófico"*. Oviedo: Pentalfa Ediciones.
- Hayes, S.C. (1984). Making sense of spirituality. *Behaviorism*, 12, 99-110.
- Luciano, M.C. y Hayes, S.C. (2001). Trastorno de evitación experiencial. *Revista Internacional de Psicología Clínica y de la Salud*, 1 (1), 109-157.
- Pérez Álvarez, M. (1999). Psicología clínica y iatrogénesis. En J. Buendía (Ed.). *Psicología clínica: Perspectivas actuales*. Madrid: Pirámide.
- Pérez Álvarez, M. (2003). *Las cuatro causas de los trastornos psicológicos*. Madrid: Editorial Universitas.
- Quiroga Romero, E. (1999). La personalidad como estructura dinámica conductual definida por la conjugación Autor-Actor. *Apuntes de Psicología*, 17 (3), 309-325.
- Watzlawick, P. (1975/1986). *El arte de amargarse la vida*. Barcelona: Editorial Heder.
- Wilde, O. (1889/2000). *La decadencia de la mentira*. Madrid: Editorial Siruela.
- Yalom, I. D. (1992/1999). *El día en que Nietzsche lloró*. Buenos Aires: Editorial Emecé.
- Yalom, I. D. (1996/2000). *Desde el diván*. Buenos Aires: Editorial Emecé.



# PERSONALIDAD Y DESEABILIDAD SOCIAL EN CONTEXTOS ORGANIZACIONALES: IMPLICACIONES PARA LA PRÁCTICA DE LA PSICOLOGÍA DEL TRABAJO Y LAS ORGANIZACIONES

## PERSONALITY AND SOCIAL DESIRABILITY IN ORGANIZATIONAL SETTINGS: PRACTICAL IMPLICATIONS FOR WORK AND ORGANIZATIONAL PSYCHOLOGY

Jesús F. Salgado

Universidad de Santiago de Compostela

Este artículo presenta una revisión de la validez de las diversas medidas de personalidad para predecir diversos criterios organizacionales, entre los que se incluyen, el desempeño laboral, el éxito en la formación, la emergencia del liderazgo, la eficacia del liderazgo, los accidentes laborales, la satisfacción laboral, la rotación en el empleo, las conductas contraproductivas, el absentismo y el salario. Seguidamente, se ocupa de examinar uno de los problemas principales a los que se han enfrentado las medidas de personalidad en el trabajo: la deseabilidad social y la distorsión de las respuestas a los cuestionarios. De este examen se desprende que la deseabilidad social tiene un efecto promedio de inflar (o desinflar) las puntuaciones alrededor de 0.38 unidades de desviación en contextos de selección, aunque no afecta a la validez predictiva de las medidas de personalidad, ni es una variable mediadora, moderadora o supresora de la validez de dichas medidas. También se observa que la deseabilidad social no está relacionada con el desempeño en el trabajo. En tercer lugar, se revisan las distintas estrategias utilizadas para reducir y neutralizar la deseabilidad y se observa que sólo dos de ellas son efectivas: (a) informar a los evaluados de que se examinarán sus respuestas en relación con la deseabilidad social y que la distorsión podrá tener consecuencias negativas para los distorsionadores, y (b) crear baremos a partir de muestras de solicitantes o de personas que contestan a las medidas de personalidad en contextos que puedan suscitar la deseabilidad social (pe., decisiones de promoción). La última parte del artículo contiene las conclusiones y las sugerencias para los profesionales de la Psicología del Trabajo y las Organizaciones.

This article consists of four sections. The first section presents a validity review of personality measures for predicting several organizational criteria, including job performance, training success, leadership emergence, leadership effectiveness, work accidents, job satisfaction, turnover, counter-productive behaviours, absenteeism, and salary. Next, the literature on social desirability and the distortion of responses to the personality questionnaires, one of the main problems of personality measures, is examined. This examination points out that social desirability has an average effect size of 0.38 standard deviation units on the personality measures, inflating the scores in personnel selection settings. However, social desirability has no effect on the predictive validity of personality measures and it is not a moderator, mediator or suppressor variable. It is also observed that social desirability is not related to job performance. In the third section, the various strategies developed for reducing social desirability are reviewed and it was found that only two are effective: (a) warning applicants that the responses will be checked for social desirability and that the distortion will be penalized, and (b) to develop norms using applicant samples or samples including individuals responding personality measures in contexts which can produce social desirability (e.g., promotion decisions). The article finishes with some conclusions and some suggestions for the practitioners of Work and Organizational Psychology.

**E**n los últimos quince años, la evaluación de la personalidad se ha convertido en un tema de gran interés para los profesionales e investigadores de la Psicología del Trabajo y las Organizaciones debido a su utilidad para tomar decisiones en relación con procesos revelantes para el trabajo. Diversos meta-análisis llevados a cabo tanto en América, como en Europa, África y Asia, han llegado esencialmente a las mismas conclusiones: las medidas de personalidad son

buenos predictores de diversos criterios organizacionales relevantes (ver Barrick y Mount, 1991; Barrick, Mount y Judge, 2001; Hogan y Holland, 2003; Hough, 1992; Hurtz y Donovan, 2000; Mount y Barrick, 1995; Rothman, Meining y Barrick, 2002; Salgado, 1997; 1998; 2002; 2003; Yoo y Ming, 2002). La base fundamental de estas conclusiones convergentes es la utilización del modelo de personalidad de los cinco grandes factores (*Big Five*) como taxonomía para integrar los resultados de los centenares de estudios de validez singulares (locales) llevados a cabo durante más de sesenta años. De acuerdo a este modelo, cinco grandes dimensiones de

Correspondencia: Jesús F. Salgado. Departamento de Psicología Social, Básica y Metodología. Universidad de Santiago de Compostela, 15782 Santiago de Compostela. España.  
E-mail: psjesal@usc.es

**TABLA 1**  
**RESUMEN DE LOS RESULTADOS META-ANALÍTICOS SOBRE LA**  
**RELACIÓN ENTRE LAS MEDIDAS DE PERSONALIDAD Y DIVERSOS**  
**CRITERIOS Y VARIABLES ORGANIZACIONALES.**  
 (Fuente original de la Tabla: Salgado y De Fruyt, 2005)

Dimensión	K	N	Validez
Desempeño Laboral <sup>a</sup>			
Conciencia	133	33,668	.33
Estabilidad Emocional	108	19,880	.21
Extroversión	111	21,916	.10
Apertura a la Experiencia	82	13,895	.09
Amigabilidad	110	21,911	.19
Éxito en la Formación <sup>b</sup>			
Conciencia	20	3,909	.31
Estabilidad Emocional	25	3,753	.09
Extroversión	21	3,484	.28
Apertura a la Experiencia	18	3,177	.33
Amigabilidad	24	4,100	.14
Emergencia del Liderazgo <sup>c</sup>			
Conciencia	17	nd	.33
Estabilidad Emocional	30	nd	.24
Extroversión	37	nd	.33
Apertura a la Experiencia	20	nd	.24
Amigabilidad	23	nd	.05
Efectividad del Liderazgo <sup>c</sup>			
Conciencia	18	nd	.16
Estabilidad Emocional	18	nd	.22
Extroversión	23	nd	.24
Apertura a la Experiencia	17	nd	.24
Amigabilidad	19	nd	.16
Satisfacción Laboral <sup>d</sup>			
Conciencia	79	21,719	.26
Estabilidad Emocional	92	24,527	.29
Extroversión	75	20,184	.25
Apertura a la Experiencia	50	15,196	.02
Amigabilidad	38	11,856	.17
Conductas Desviadas (Contraproductivas) (Invertida la Validez) <sup>e</sup>			
Conciencia	13	6,276	.26
Estabilidad Emocional	15	3,107	.06
Extroversión	12	2,383	-.01
Apertura a la Experiencia	8	1,421	-.14
Amigabilidad	9	1,299	.20
Rotación en el Empleo (Invertida la Validez) <sup>e</sup>			
Conciencia	5	748	.31
Estabilidad Emocional	4	554	.35
Extroversión	4	554	.20
Apertura a la Experiencia	4	554	.14
Amigabilidad	4	554	.22
Accidentes Laborales (Invertida la Validez) <sup>f</sup>			
Conciencia	9	1125	.30
Estabilidad Emocional	13	1198	.28
Extroversión	12	1524	-.09
Apertura a la Experiencia	7	570	-.50
Amigabilidad	7	420	.61

Nota. K= número de estudios; N= Tamaño total de muestra; nd= no disponible; <sup>a</sup>=Salgado (2004); <sup>b</sup> Barrick, Mount y Judge (2001); <sup>c</sup> Judge, Bono, Ilies, y Gerhardt (2002); <sup>d</sup> Judge, Heller y Mount (2002); <sup>e</sup> Salgado (2002); <sup>f</sup> Clarke y Robertson (2005).

personalidad han resultados replicables a través de diferentes muestras, en diferentes culturas, con diferentes lenguajes y diferentes técnicas evaluativas. La denominación de tales factores varía entre los diversos investigadores del campo de la psicología de la personalidad en el trabajo, pero la más ampliamente utilizada es la sugerida por Costa y McCrae (1992). De acuerdo a ellos, las cinco dimensiones o factores de la personalidad serían: estabilidad emocional (reverso de neuroticismo), extraversión (opuesto a introversión), apertura a la experiencia (frente a cierre a la experiencia), amigabilidad (opuesto a antagonismo) y conciencia (frente a irresponsabilidad). Las denominaciones de los factores de estabilidad emocional y extraversión son las que más consenso han obtenido. Las tres restantes, quizás por ser más recientes, han recibido otras denominaciones alternativas. Por ejemplo, la dimensión de apertura a la experiencia ha sido denominada también cultura o intelectualidad; la dimensión de amigabilidad ha sido también llamada amistosidad y la dimensión de conciencia ha recibido etiquetas tales como escrupulosidad, minuciosidad, responsabilidad o afán de logro. Los resultados de los meta-análisis mencionados más arriba han demostrado que dos factores de personalidad, estabilidad emocional y conciencia, son predictores válidos del desempeño ocupacional en todas las ocupaciones. También se ha demostrado que estos dos factores junto con extraversión son predictores válidos de éxito en la formación (*training proficiency*) y que el factor de conciencia y el de amigabilidad predicen el comportamiento contraproducente.

Además de lo anterior, la investigación en el ámbito de la personalidad en las organizaciones también ha demostrado que las medidas de la personalidad centradas en criterios ocupacionales (*Criterion-focused Occupational Personality Scales, COPS*), tales como los tests de integridad, orientación al servicio del cliente, potencial de dirección, tolerancia al stress o potencial comercial son excelentes predictores de diversos criterios organizacionales, incluyendo el desempeño ocupacional, el éxito en la formación y los comportamientos contraproducentes (ver Ones y Viswesvaran, 2001a y b, para un resumen). La investigación también ha demostrado que estas medidas (COPS) son una combinación de tres de los factores básicos de la personalidad: estabilidad emocional, amigabilidad y conciencia. En las Tablas 1, 2 y 3 que aparecen a continuación, figura un resumen de la validez de

las distintas dimensiones y compuestos de personalidad para la predicción de diversos criterios organizacionales.

A la vista de los resultados de las investigaciones meta-analíticas mencionadas, resulta evidente por qué las medidas de personalidad han sido utilizadas profusamente en las decisiones organizacionales en la última década y que hayan suscitado el interés de los profesionales. Aunque de modo prioritario se han utilizado para propósitos de selección de personal, también se han utilizado en procesos de formación, de desarrollo (p.e., *coaching*), o para el establecimiento de perfiles de competencias.

Sin embargo y a pesar de este reciente éxito de las medidas de personalidad para predecir el desempeño ocupacional, su utilización en ámbitos organizacionales no está exenta de problemas y dificultades. Entre ellos, la distorsión y el sesgo en las respuestas a los ítems que componen dichas medidas han sido dos de los que han recibido mayor interés y sobre los que ha habido mayor preocupación, resultando, en consecuencia, abundantes los estudios que se han ocupado de esta cuestión en los últimos años.

Hace sesenta años ya que Mehl y Hathaway (1946) y Ellis (1946) señalaron que, en efecto, las personas cuando se les instruye para hacerlo pueden distorsionar las respuestas a los cuestionarios de personalidad

**LA DISTORSIÓN DE LAS RESPUESTAS A LOS ÍTEMS DE LAS MEDIDAS DE PERSONALIDAD**

La distorsión y el sesgo en las respuestas a los ítems de las medidas de personalidad, especialmente en contextos organizacionales, donde las decisiones basadas en las respuestas a los cuestionarios de personalidad tienen importantes implicaciones para las personas que contes-

**TABLA 2**  
**RESUMEN DE LOS RESULTADOS META-ANALÍTICOS SOBRE LA RELACIÓN ENTRE COPS Y DIVERSOS CRITERIOS Y VARIABLES ORGANIZACIONALES.**  
(Fuente original de la tabla: Salgado y De Fruyt, 2005)

Dimensión	K	N	Validez
Desempeño Laboral			
Integridad (tests de personalidad) <sup>a</sup>	102	27,081	.37
Escalas de Drogas y Alcohol <sup>b</sup>	7	1,436	.19
Escalas de Tolerancia al Stress <sup>b</sup>	13	1,010	.42
Escalas de Servicio al Cliente <sup>b</sup>	33	6,944	.39
Escalas de Violencia <sup>c</sup>	14	4,003	.41
Éxito en la Formación			
Integridad (tests de personalidad) <sup>d</sup>		2,364	.38
Conductas Desviadas (Contraproductivas) (Invertida la Validez) <sup>e</sup>			
Integridad (tests de personalidad) <sup>a</sup>	138	158,065	.32
Escalas de Tolerancia al Stress <sup>b</sup>	5	594	.42
Escalas de Servicio al Cliente <sup>b</sup>	5	740	.42
Escalas de Violencia <sup>c</sup>	4	533	.46
Absentismo			
Integridad (tests de personalidad) <sup>e</sup>	16	5,435	.36

Nota. K= numero de estudios; N= Tamaño total de la muestra; <sup>a</sup> =Ones, Viswesvaran y Schmidt (1993); <sup>b</sup> = Ones y Viswesvaran (2001a); <sup>c</sup> = Ones y Viswesvaran (2001b); <sup>d</sup>= Ones y Viswesvaran (1998a); <sup>e</sup>= Ones, Viswesvaran y Schmidt (2003).

**TABLA 3**  
**RESUMEN DE LOS RESULTADOS META-ANALÍTICOS ENTRE VARIAS VARIABLES DE PERSONALIDAD (Y MODELOS) Y DIVERSOS CRITERIOS Y VARIABLES ORGANIZACIONALES.**  
(Fuente original de la tabla: Salgado y De Fruyt, 2005)

Dimensión	K	N	Validez
Desempeño Laboral			
Conciencia-NFFM <sup>a</sup>	36	5,874	.18
Estabilidad Emotional-NFFM <sup>a</sup>	25	4,541	.05
Extroversión-NFFM <sup>a</sup>	26	4,338	.08
Apertura a la Experiencia-NFFM <sup>a</sup>	29	4,364	.08
Amigabilidad-NFFM <sup>a</sup>	31	4,573	.13
Autoeficacia Generalizada <sup>b</sup>	11	1,506	.43
Locus de Control <sup>c</sup>	35	4,310	.22
Autoestima <sup>c</sup>	40	5,145	.26
Éxito en la Formación			
Autoeficacia Generalizada <sup>b</sup>	4	422	.29
Rendimiento en el Puesto			
Inteligencia Emocional <sup>d</sup>	19	2,652	.24
Satisfacción Laboral			
Afectividad Positiva <sup>e</sup>	15	3,326	.49
Afectividad Negativa <sup>e</sup>	27	6,233	-.33
Autoeficacia Generalizada <sup>c</sup>	8	1,411	.29
Locus de Control <sup>c</sup>	80	18,491	.32
Autoestima <sup>c</sup>	56	20,819	.26
Salario			
Autoeficacia Generalizada <sup>b</sup>	5	468	.28
Absentismo			
Autoeficacia Generalizada <sup>b</sup>	4	718	.21

Nota. K= número de estudios; N= Tamaño total de la muestra; NFFM= medidas a partir de cuestionarios no basados en el modelo de los cinco grandes factores; <sup>a</sup> =Salgado (2003); <sup>b</sup>=Salgado y Moscoso (2000); <sup>c</sup> =Judge y Bono (2001) <sup>d</sup>=Van Rooy y Viswesvaran (2004); <sup>e</sup> =Connolly y Viswesvaran (2000).

tan a dichos cuestionarios (pe., ser contratado o no), ha recibido considerable atención desde hace más de cincuenta años. La distorsión en las respuestas puede ser tanto de signo positivo (pe., tratar de causar una buena impresión), como de signo negativo (pe., tratar de sugerir que hay un problema o trastorno psicológico). En el ámbito de las organizaciones, aunque el segundo tipo de sesgo existe en determinadas circunstancias, es el primero, la distorsión positiva, la que ha resultado más preocupante, dado que las medidas de personalidad han sido utilizadas preferentemente con propósitos de selección de personal y, por tanto, 'tratar de causar una buena impresión' es una opción frecuente como respuesta a los tests y cuestionarios. A la vista de que la distorsión de las respuestas en un sentido favorable podría favorecer las posibilidades de los candidatos para ser contratados, Seisdedos (1988), en uno de los escasos estudios realizados en España sobre este fenómeno en las organizaciones, ha denominado a este sesgo "Adaptación Inteligente". Sin embargo, las denominaciones frecuentes para este sesgo son otras, tales como 'Deseabilidad Social', 'Sinceridad' y 'Distorsión Motivacional'. También se han utilizado denominaciones como 'Simulación', 'Virtudes Infrecuentes' o 'Distorsión de la Respuesta'. De todas estas denominaciones, 'Deseabilidad Social' es la más frecuentemente utilizada y todas las demás caben dentro de esta etiqueta.

La deseabilidad social o distorsión de la respuesta ha sido definida como "la tendencia a contestar a los ítems de modo que se responde a las presiones sociales o normativas en lugar de proporcionar un autoinforme verídico" (Ellingson, Smith y Sackett, 2001, p.122). Aunque muchas veces la deseabilidad social ha sido considerada como un constructo unidimensional y muchas medidas sólo proporcionan una medida global de la misma, de acuerdo a la investigación más reciente, llevada a cabo principalmente por Paulhus (1984, 2002), la deseabilidad consiste en dos dimensiones, que han sido denominadas manejo de impresiones (*impresión management*) y auto-engaño (*self-deception*). El manejo de impresiones indica una tendencia a adaptar intencionalmente la imagen pública de uno mismo con objeto de obtener una visión favorable por parte de los demás. El auto-engaño, por el contrario, se refiere a la tendencia no intencionada de describirse a uno mismo de un modo favorable y que se manifiesta en auto-descripciones ses-

gadas positivamente pero en las que se cree honestamente. En este sentido, el manejo de impresiones es una manipulación voluntaria de la propia imagen para que los demás nos perciban de modo positivo mientras que el auto-engaño no es una manipulación deliberada aunque pueda conducir a distorsiones en la percepción que los otros tienen de nosotros. Teniendo en cuenta esta diferenciación entre las dos dimensiones de la deseabilidad social, en el ámbito de la Psicología del Trabajo y las Organizaciones, los efectos del manejo de impresiones sobre las puntuaciones de las medidas de personalidad parecen más relevantes.

La diferenciación entre manejo de impresiones y auto-engaño es relevante a la vista de lo sugerido por algunos investigadores del campo de la psicología de la personalidad y según lo cual existen diferencias individuales en la deseabilidad social (Block 1965; McCrae y Costa, 1983). Esto significa que la deseabilidad social podría no ser sólo una tendencia a adaptarse a las situaciones sino una característica estable de la personalidad que indicaría diferencias más sustantivas y significativas. Por ejemplo McCrae y Costa, (1983), Ones, Viswesvaran y Reiss (1996) y Salgado, Iglesias y Remeseiro (1996) han encontrado que la deseabilidad social correlaciona con la estabilidad emocional y con la dimensión de conciencia. Como han señalado McCrae y Costa (1983), esto significaría que una persona que verdaderamente puntúa alto en el factor de conciencia, que es emocionalmente estable y cooperativo (puntúa alto en amigabilidad) aparecería también como alto en deseabilidad social. Sin embargo, y esta es la paradoja, sería una persona honesta y fiable pero sería acusada de distorsionar o mentir en los cuestionarios de personalidad.

Los efectos de la deseabilidad social sobre las medidas de la personalidad y sobre otros instrumentos evaluativos (pe., entrevistas) están ampliamente documentados en la literatura y son particularmente conocidos por los responsables de la selección de personal en las organizaciones. El principal efecto es que la deseabilidad social tiende a inflar (aumentar) las puntuaciones en las dimensiones que están (o que los candidatos creen que están) relacionadas positivamente con el desempeño laboral y a desinflar (reducir) las puntuaciones en aquellas dimensiones que están (o que los candidatos creen que están) relacionadas negativamente con el desempeño laboral. La investigación reciente ha revisado meta-analíticamen-

te y mediante estudios con muestras grandes los efectos de la deseabilidad social sobre las respuestas a las medidas de personalidad. En general, tal investigación empírica ha consistido en la utilización de tres tipos de diseños: (a) comparaciones entre grupos en situaciones de laboratorio, con participantes instruidos para distorsionar sus respuestas; (b) comparaciones intra-sujetos en situaciones de laboratorio, con participantes instruidos para distorsionar, y (c) comparaciones en situaciones reales de selección, examinando las diferencias entre las personas que muestran deseabilidad social y las que no.

Ones y sus colaboradores (Ones y Viswesvaran 1998a y b ; Ones, Viswesvaran y Reiss, 1996; Viswesvaran, Ones y Hough, 2001), Hough y sus colaboradores (1998; Hough, Eaton, Dunnette, Kamp y McCloy, 1990; Hough y Paullin, 1994) y Christiansen y sus colaboradores (Christiansen, Goffin, Johnston y Rothstein, 1994; Goffin y Christiansen, 2003) son los principales investigadores de los efectos de la deseabilidad social en ámbitos organizacionales, aunque muchos otros investigadores han hecho contribuciones relevantes. Los principales resultados han sido los siguientes. En los estudios con comparaciones entre grupos en situaciones de laboratorio con distorsión positiva inducida se encontró que, en los cuestionarios que miden dimensiones de personalidad (pe., los *Big Five*) las personas distorsionadoras puntúan, en promedio, 0,6 unidades de desviación más que las no distorsionadoras. Esto significa lo siguiente: si utilizamos puntuaciones *T* (media=50; SD=10), el grupo de personas distorsionadoras puntuaría en promedio 56 frente a una puntuación de 50 correspondiente al grupo de personas no distorsionadoras. Cuando los estudios realizan comparaciones intra-sujeto en situaciones de laboratorio con distorsión inducida y con respuesta honesta, la diferencia en las puntuaciones en las medidas de las dimensiones de personalidad entre las dos condiciones para los mismos sujetos es de 0,72 unidades de desviación típica, lo que en términos de puntuaciones *T* significa que la distorsión hace elevar las puntuaciones de una misma persona de 50 a 57,2 puntos. Cuando no se tiene en cuenta el tipo de diseño, la diferencia es de 0,5 unidades de desviación. Por tanto, estos resultados indican que la deseabilidad social tiene efectos sobre las puntuaciones de las personas distorsionadoras que se sitúan ligeramente por encima de media unidad de desviación. También muestran que el tipo de diseño de investigación tiene efectos importantes sobre el

tamaño de la distorsión. Este mismo tipo de estudios se realizaron con las medidas de compuestos de personalidad (COPS), por ejemplo los tests de integridad, y los resultados han sido muy semejantes ya que la diferencia entre las personas instruidas para contestar de un modo distorsionado positivo frente a las personas instruidas para contestar honestamente ha sido de 0,50 unidades de desviación.

Los resultados acabados de exponer pueden considerarse indicadores de los efectos de la deseabilidad social en situaciones de rendimiento máximo (cuando trata de encontrarse la mayor diferencia posible entre las condiciones de deseabilidad y de honestidad). Sin embargo, los estudios sobre los efectos de la deseabilidad social en condiciones "reales", esto es, cuando se comparan las respuestas de las personas distorsionadoras y las no distorsionadoras en situaciones verdaderas de selección, serían indicadores de situación de rendimiento típico (cuando se trata de encontrar la diferencia habitual entre las dos condiciones mencionadas en un contexto típico laboral). A este respecto, Hough (1998) llevó a cabo tres estudios particularmente relevantes. En el primero de ellos comparó una muestra de 963 ocupantes de puestos en la industria de las telecomunicaciones con una muestra de 14.442 solicitantes de empleo en la misma industria y encontró que, en una medida de responsabilidad (una de las subdimensiones del factor de conciencia), existía una distorsión promedio de 0,45 unidades de desviación a favor de los solicitantes. En un segundo estudio, en puestos para la policía local, con una muestra de ocupantes de 508 personas y otra de solicitantes de 24.433, usando una medida del factor de conciencia, encontró que la distorsión promedio era de 0,33 unidades de desviación a favor de los solicitantes. En el tercer estudio, con una muestra de 270 guardias estatales y otra de solicitantes de 681 personas, encontró una distorsión promedio de 0,13 en una escala de personalidad compuesta de las facetas de búsqueda de experiencia, y auto-estima. Otros estudios llevados a cabo por Hough, Eaton, Dunnette, Kamp y McCloy (1990), reflejan resultados semejantes. Por tanto, en conjunto los datos disponibles indican que en el tercer tipo de diseño de investigación (comparación de ocupantes y solicitantes en situaciones reales), los efectos promedio de la deseabilidad social, aunque existen, son mucho menores que los encontrados en las situaciones de laboratorio. En otras palabras, en situaciones de rendimiento típico (pe.,



selección de personal), los efectos de la deseabilidad social se reducen en comparación con los encontrados en situaciones de rendimiento máximo (pe., manipulación experimental).

Además del efecto de inflar las puntuaciones en las medidas de personalidad, un segundo posible efecto que ha sido mencionado en diversas ocasiones y que, probablemente, es el que más ha preocupado a los profesionales de la Psicología del Trabajo y las Organizaciones que tienen responsabilidades derivadas del uso de las medidas de medidas de personalidad (pe., decisiones de contratación, formación desarrollo, promoción, etc.), es el que se refiere al impacto de la deseabilidad sobre la validez de dichas medidas. Algunos investigadores y no pocos profesionales creen que la deseabilidad social afecta negativamente a la validez de las medidas, haciendo que esta se reduzca y que, en consecuencia, lleguen a ser inválidas. Dicho en otras palabras, la capacidad de predicción del desempeño ocupacional y otros criterios organizacionales que tienen las medidas de personalidad se vería sustancialmente reducida o incluso anulada como consecuencia de los efectos de la deseabilidad (Goffin y Christiansen, 2003; Mueller-Hanson, Heggstad y Thornton, 2003; Rosse, Stecher, Miller y Levin, 1998). Esta creencia ha llevado a que se haya cuestionado seriamente el uso de las medidas de personalidad en ámbitos organizacionales y que incluso algunos profesionales rechacen abiertamente su utilidad para tomar decisiones.

A este respecto, se han llevado a cabo recientemente investigaciones meta-analíticas y estudios singulares con muestras grandes ( $N > 1200$ ) que han examinado los efectos de la deseabilidad social sobre la validez de las medidas de personalidad. Así, por ejemplo, en relación con la validez de constructo, Ones, Viswesvaran y Reiss (1996) han demostrado que la deseabilidad social apenas tiene efecto sobre la validez convergente y discriminante de las medidas de personalidad, siendo los crementos promedio en las correlaciones del orden de .015, es decir, irrelevantes desde el punto de vista práctico. Por lo que respecta a la validez de criterio (predictiva), Ones y Viswesvaran (1998b) han examinado sus efectos tanto si se la considera como una variable moderadora, mediadora o supresora. Los resultados de sus meta-análisis indican que la deseabilidad social no tiene impacto sobre la validez de las medidas de personalidad, tanto sean referidas a las dimensiones básicas (pe.,

*Big Five*), como a compuestos (pe., integridad), permaneciendo la validez esencialmente la misma una vez que han sido parcializados los efectos de la deseabilidad social. Los estudios de Hough (1998; Hough y otros, 1990) llegaron básicamente a los mismos resultados. En consecuencia, se demostró que era falsa la creencia sostenida por algunos investigadores y bastantes profesionales respecto a los efectos negativos de la deseabilidad social sobre la validez de las medidas de personalidad.

### ¿COMO SE DETECTA (MIDE) LA DESEABILIDAD SOCIAL?

Dado el interés que ha suscitado la cuestión de la deseabilidad social y sus efectos sobre las puntuaciones de las medidas de personalidad, diversos investigadores se plantearon como detectarla y, a lo largo de los años, diferentes instrumentos han sido creados, aunque casi todos ellos beben en la fuente común del *Minnesota Multiphasic Personality Inventory* (MMPI). En efecto, los creadores del MMPI, Hathaway y MacKinley, desarrollaron dos escalas para detectar la posible distorsión en las respuestas al cuestionario. Una es llamada la escala K y con ella se trata de detectar los sesgos negativos o, lo que es lo mismo, la tendencia a aparentar una peor imagen de uno mismo. La segunda es la escala *Lie* (mentira) que en ocasiones se interpreta como sinceridad y que trata de detectar los sesgos positivos o la tendencia a aparentar una buena imagen personal. Ambas escalas fueron desarrolladas como tests o pruebas orientadas al criterio, como lo habían sido las restantes escalas del MMPI. Siguiendo la senda del MMPI, Harrison Gough, el creador del *California Personality Inventory* (CPI; 1987) y alumno de Hathaway, desarrolló también una escala para medir la deseabilidad social en su cuestionario y la denominó *Good Impression Scale* (Escala de Buena Impresión). Puesto que el CPI estaba diseñado para evaluar la personalidad normal (ajustada) a diferencia del MMPI que se había diseñado para evaluar desórdenes de personalidad (desajuste), el interés de Gough residía en la tendencia a distorsionar positivamente. Eysenck, en su primer cuestionario de personalidad, *Maudsley Personality Inventory* (MPI; predecesor del *Eysenck Personality Inventory*; EPI), incluyó también una escala para evaluar la deseabilidad social. Dicha escala fue derivada también de la escala L del MMPI. Sobre esta senda, a lo largo de los años, se han creado multitud de escalas para medir la deseabilidad social y muchos cuestionarios de perso-

nalidad incluyen entre sus ítems alguna escala o medida de esta distorsión. Además de las mencionadas más arriba, han sido muy populares, por ejemplo, la escala de deseabilidad social de Edwards (1957), la escala de deseabilidad social de Crowne-Marlowe (1964), la escala de sinceridad de Eysenck (Eysenck y Eysenck, 1964; originalmente es una escala de mentira), la escala de distorsión motivacional positiva del 16PF (Cattell, Eber y Tatsuoka, 1970) y la escala de deseabilidad social de Occupational Personality Questionnaire (SHL, 1999). La totalidad de estas escalas fueron confeccionadas considerando que la deseabilidad social es un concepto unidimensional. Sin embargo, a raíz de las investigaciones de Paulhus ya mencionadas, hoy se acepta que la deseabilidad social consiste en dos dimensiones y el *Balanced Inventory of Desirable Responding* (BIDR, Paulhus 1984, 2002) es el más popular de los inventarios modernos. Este inventario contiene 40 ítems, con dos subescalas de 20 ítems cada una, que evalúan el 'manejo de impresiones' y el 'auto-engaño'.

No obstante, es de señalar también que en la actualidad muchos de los cuestionarios recientes de personalidad, sobre todo los basados en el modelo de los cinco grandes factores, no utilizan ninguna escala para detectar la deseabilidad social. Este es el caso, por ejemplo, del NEO-PI-R (Costa y McCrae, 1992), el *Hogan Personality Inventory* (HPI, Hogan y Hogan, 1995), el IP/5F (Salgado, 1996) y el *Jackson Personality Inventory* (JPI, Jackson, 1994).

### ESTRATEGIAS PARA REDUCIR LOS EFECTOS DE LA DESEABILIDAD SOCIAL

Una vez determinados los efectos de la deseabilidad social en contextos organizacionales y viendo que éstos se reducen a la inflación o deflación de las puntuaciones de las medidas de algunas dimensiones y facetas de la personalidad, la siguiente cuestión a tratar es cómo se pueden reducir o anular dichos efectos.

A este respecto, a lo largo de los años, se han propuesto diversas estrategias de muy distinta naturaleza. Entre las mismas figuran las siguientes: (a) utilización de escalas con ítems de elección forzosa, habiendo emparejado los ítems de acuerdo a su similitud en deseabilidad social (esta modalidad se denomina también 'medida ipsativa' o 'puntuación ipsativa'), (b) usar escalas de detección de la deseabilidad y descartar a aquellas personas que

puntúan moderadamente alto (pe., 2 unidades de desviación sobre la media) en dichas escalas; (c) usar las escalas de detección y subjetivamente "ajustar" las puntuaciones en las medidas de personalidad de aquellas personas que puntúan moderadamente alto en las escalas de deseabilidad; (d) usar escalas de detección y "corregir" mecánicamente las puntuaciones en las medidas de personalidad de aquellas personas que distorsionan moderadamente alto y utilizando para ello una fórmula matemática; (e) informar a los evaluados sobre la existencia de métodos de detección durante la evaluación e instruirlos sobre las posibles consecuencias de la distorsión; (f) desarrollar baremos específicos para muestras de solicitantes en vez de utilizar los baremos de las muestras normativas de la población general. En la Tabla 4 figura un resumen de las estrategias anteriores

En la actualidad es realmente escasa la utilización de escalas con formato de elección forzosa para evaluar la personalidad en ambientes organizacionales y, especialmente, para propósitos de selección de personal. No obstante, existen cuestionarios que emplean este formato con la idea de que reducirá la distorsión de las respuestas. Ejemplos de cuestionarios de personalidad que utilizan este formato de respuesta son el *Occupational Personality Questionnaire 3.2* (OPQ 3.2; SHL, 1999), el *Personality Inventory* de Thomas, también conocido como el DISC (Thomas Internacional) o el *Description in Cinq Dimensions* (D5D, Rolland y Mogenet, 2001). El supuesto de partida de los creadores de cuestionarios con este formato es que si los ítems están agrupados, por ejemplo en tétradas, con similar deseabilidad social y los evaluados tienen que indicar cual de los ítems es el que mejor lo define y cuál es el que peor, descartando los otros dos, la respuesta final reflejará mejor sus características de personalidad y eliminará los efectos de la deseabilidad social. Hicks (1970, p.181) sugirió que la justificación para el uso de medidas de elección forzosa (ipsativas) requeriría demostrar tres condiciones: (1) que existe un sesgo importante en las respuestas a los cuestionarios de personalidad, (2) que este sesgo disminuye la validez y (3) que el formato de elección forzosa disminuye el sesgo y aumenta la validez en mayor cantidad que lo hacen otros controles no ipsativos del sesgo. Este autor concluyó que no se había dado ningún caso en el que se demostrasen esas tres condiciones conjuntamente. Veinticinco años más tarde, Bartram (1996) consideraba

que lo conclusión de Hicks seguía siendo verdad. A este respecto hay que indicar lo siguiente: (a) la investigación más reciente y exhaustiva, mencionada en apartados precedentes, ha demostrado que existe un sesgo y que puede ser importante en situaciones reales de evaluación en contextos organizacionales (pe., selección de personal) y (b) la investigación también ha demostrado que la deseabilidad social no disminuye la validez. Con respecto a la tercera condición, la referida a que los formatos de elección forzosa (ipsativos) disminuyen el sesgo y aumentan la validez, en los últimos años se han publicado diversas investigaciones que examinaron esta cuestión. Por ejemplo, Christiansen, Burns y Montgomery (2005) han realizado un conjunto de estudios en los que muestran que (1) las escalas de elección forzosa son tan susceptibles de ser distorsionadas como las escalas tradicionales normativas, (2) que las personas con pun-

tuciones más altas en las medidas de habilidad mental general tienen más éxito en mejorar sus puntuaciones en las escalas de elección forzosa (en el sentido de ser más ajustados al perfil deseable), sugiriendo, por tanto, que tales escalas son manipulables en función de la habilidad mental general de los evaluados. Baron (1996), una ardiente defensora de este formato para las medidas de personalidad, reconoce que un pequeño grupo de candidatos distorsiona de un modo potente las puntuaciones en estos cuestionarios. En consecuencia, la tercera condición planteada por Hicks no se cumple en la práctica a la luz de los datos recientes. Por tanto, y teniendo en cuenta que este formato de respuesta tiene muchas y serias limitaciones de índole psicométrica, tanto en lo que se refiere a su fiabilidad, como a su factorización, como a su validez y a la comparación entre individuos (véanse, por ejemplo, las revisiones de Baron, 1996; Bartram, 1996;

**TABLA 4**  
**ESTRATEGIAS DE REDUCCIÓN DE LA DESEABILIDAD SOCIAL Y SUGERENCIAS A LOS PROFESIONALES**

Estrategia	Descripción	Limitaciones	Eficacia	Recomendación
Escalas de Elección Forzosa	Elegir entre ítems con similar grado de deseabilidad social	Teóricas, metodológicas y prácticas	Limitada	No recomendable
Usar escalas de Deseabilidad Social y descartar a los sujetos altos en DS.	Incluir una medida de distorsión y excluir a los que puntúen por encima de un punto de corte.	Teóricas, prácticas y legales	Ineficaz	No recomendable
Usar escalas de Deseabilidad Social y ajustar las puntuaciones de los distorsionadores (Estrategia Subjetiva)	Se ajustan "subjetivamente" las puntuaciones de los sujetos que se consideran distorsionadores, basándose en la experiencia del evaluador.	Teóricas y prácticas	Ineficaz	No recomendable
Usar escalas de Deseabilidad Social y ajustar las puntuaciones de los distorsionadores (Estrategia objetiva)	Se ajustan "mecánicamente" las puntuaciones de los sujetos que se consideran distorsionadores, basándose, por ejemplo, en una ecuación de regresión.	Teóricas y prácticas	Ineficaz	No recomendable
Informar a los evaluados	Se informa a los evaluados que existe la posibilidad de ser eliminados o penalizados si distorsionan sus respuestas.	No tiene	Eficaz	Recomendable
Elaborar baremos basados en muestras de solicitantes	Baremar las puntuaciones de los evaluados después de crear un baremo confeccionado con muestras de solicitantes en vez de una muestra de la población en general.	No tiene	Eficaz	Recomendable

Closs, 1996; Cornwell y Dunlap, 1994; Hicks, 1970 y Meade, 2004), la decisión respecto a esta estrategia es que debería ser descartada y no utilizada con propósitos de evaluación en contextos organizacionales en los cuáles haya que comparar personas entre sí (pe., selección).

La segunda estrategia, descartar a aquellas personas que puntúan alto en las escalas de deseabilidad social introduce serias complicaciones de orden teórico, práctico e incluso podrían ser de orden legal. Desde el punto de vista teórico, es posible, como han señalado McCrae y Costa (1983) que personas totalmente honestas en sus respuestas, pero con unas características de personalidad que se ajustan al perfil típico del distorsionador, sean excluidas sin más consideración. Es decir, estaríamos ante casos de 'falso positivos' y esta estrategia no los consideraría. Además, y desde el punto de vista teórico, habrá que demostrar que las personas que puntúan más alto en las escalas de distorsión motivacional muestran posteriormente un desempeño laboral inferior al de las personas que no distorsionan. Y esto no sólo no se ha demostrado, sino que se ha demostrado que no hay relación entre la deseabilidad y el desempeño laboral (Ones y Viswesvaran, 1998b). Por otra parte, tampoco se ha demostrado que las personas que distorsionan y, consecuentemente, presentan puntuaciones más altas en algunas dimensiones de personalidad (pe., conciencia) relacionadas con el desempeño en el trabajo, muestren posteriormente cuando han sido contratadas, un desempeño inferior al de las personas que puntúan en niveles semejantes pero que no han distorsionado sus respuestas. Desde el punto de vista práctico, la exclusión de un número de personas del conjunto de evaluados puede ser contraproducente si el ratio de selección es alto (por ejemplo se aproxima a 1), esto es, si el número de vacantes y de candidatos es semejante. Por último, desde el punto de vista legal, la exclusión de candidatos basándose en sus respuestas distorsionadas no puede ser defendible ante los tribunales cuando se sabe, como ahora sabemos, que no hay relación negativa entre deseabilidad social y desempeño laboral. Por tanto, haber distorsionado no puede ser causa de exclusión desde un punto de vista legal, y cualquier candidato que recurriese ante los tribunales una decisión basada en este punto tendría muchas posibilidades de obtener una sentencia favorable. En consecuencia, esta estrategia tampoco debería ser utilizada por los profesionales de la Psicología

del Trabajo y las Organizaciones en determinadas prácticas en contextos laborales.

La tercera y la cuarta estrategias consisten en usar las escalas de detección y subjetivamente "ajustar" las puntuaciones en las medidas de personalidad de aquellas personas que puntúan moderadamente alto en las escalas de deseabilidad o bien "corregir" mecánicamente las puntuaciones en las medidas de personalidad de aquellas personas que distorsionan moderadamente alto y utilizando para ello una fórmula matemática. Como puede apreciarse, ambas estrategias son muy semejantes y tienen problemas parecidos. La corrección de las puntuaciones, aumentando o disminuyendo la puntuación de las personas, según el grado de distorsión detectado en la escala de deseabilidad social ha sido una estrategia muy habitual entre los psicólogos en general y entre los psicólogos del trabajo y las organizaciones en particular. Ello se debe a que dos de los más utilizados y populares cuestionarios de personalidad incluían sistemas de 'corrección' de las puntuaciones: el MMPI y el 16PF. En el caso de estos dos cuestionarios, la estrategia es la cuarta, es decir, la corrección mecánica (a partir de una ecuación de regresión derivada de la distorsión motivacional), pero ha dado lugar a que muchos profesionales considerasen la corrección de un modo subjetivo, basado en la experiencia o en el 'ojo clínico' del evaluador y ajustasen las puntuaciones de acuerdo a dicha experiencia. Un problema común a los dos métodos, y habitualmente no considerado por los profesionales, es que la corrección de las puntuaciones altera la validez de constructo de los cuestionarios (véase, Ones y Viswesvaran, 1998a y b) y, por tanto, implica la alteración de la medida y su fiabilidad, no mejorando su validez de criterio (predictiva) y sí pudiéndola empeorar (véase, Hough 1998). Por tanto, las puntuaciones derivadas de la corrección y ajuste de los datos directos podrían no responder a las auténticas características de personalidad del evaluado. Además, la estrategia tercera es prácticamente inviable cuando hay que tratar con un número elevado de evaluados ya que requiere el examen de cada perfil particular para llegar a hacer la corrección. En consecuencia, también estas dos estrategias son contrarias a una buena práctica profesional basada en el conocimiento empírico disponible hasta el momento. No es de extrañar, por tanto, que en la última versión del 16PF la corrección las puntuaciones haya dejado de utilizarse.

La quinta estrategia mencionada más arriba consiste en informar a las personas que están siendo evaluadas mediante cuestionarios de personalidad que entre los mismos se incluyen métodos de detección de la deseabilidad social y otros posibles sesgos, solicitando que sean lo más sinceros y honestos posible. Al propio tiempo se les instruye que, aquellas personas que sean detectadas como distorsionadas, podrían ser excluidas de posterior consideración o cualquier otra penalización o consecuencia negativa que se considere apropiada para el caso. Por ejemplo, Hough (1998) utilizó esta estrategia en varios procesos de selección e instruyó a los solicitantes diciéndoles que aquellas personas que distorsionaran sus puntuaciones en el cuestionario de personalidad serían detectadas y los que fuesen identificados por haber proporcionado auto-descripciones excesivamente virtuosas serían descalificadas del proceso de selección durante seis meses, momento en el que podrían ser evaluados otra vez. Aunque Hough no estimó la eficacia de esta estrategia ya que ella estaba interesada en comparar la diferencia de sesgo entre los ocupantes y los solicitantes, a partir de los datos proporcionados por Hough (1998) y los proporcionados por Ones, Viswesvaran y Reiss (1996) podemos establecer el grado de eficacia de esta quinta estrategia. Los tres estudios de Hough (1998) tienen una muestra total acumulada de 40.297 personas y la distorsión promedio ponderada por el tamaño de la muestra de los solicitantes frente a los ocupantes es de 0,37 unidades de desviación. Teniendo en cuenta que en el presente caso se trata de comparar a los distorsionadores frente a los no distorsionadores, se puede utilizar como estimación de esta última comparación la obtenida por Ones et al. (1996) para comparaciones entre grupos en situaciones de rendimiento máximo y que resultó ser de 0,60 unidades de desviación. Por tanto, descontado 0,37 unidades correspondientes a los distorsionadores en situaciones de selección de 0,60 unidades que es la distorsión máxima alcanzable, se obtiene un resultado de 0,23 unidades de desviación menos, que son directamente atribuibles a la estrategia de información a los evaluados. Así pues, los resultados de estos dos estudios indican que esta estrategia es eficaz en la reducción de la deseabilidad social, consiguiendo reducir la misma en más de un 38%. Dado que la penalización sugerida por Hough sólo postpone la decisión de contratación de los candidatos distorsionadores, esta estrategia es legalmente defendible. Más recientemente, McFarland (2003) examinó, en un contexto

de laboratorio, los efectos de esta estrategia para reducir la distorsión y evaluó sus efectos sobre las reacciones de los solicitantes en relación con la justicia organizacional percibida. Los resultados de McFarland indicaron que la estrategia era muy eficaz para reducir la deseabilidad social (0,45 unidades de desviación en promedio) y que, además, no tiene efectos negativos sobre las percepciones de los solicitantes (en un caso incluso fueron positivos) y reduce la multicolinealidad entre las variables de personalidad. En consecuencia, esta es una estrategia válida (reduce la distorsión) y económica, ya que no es necesario incluir una medida de la deseabilidad social, basta con informar que existe y que puede ocasionar una penalización.

La última estrategia mencionada consiste en desarrollar baremos específicos para contextos de selección de personal. En otras palabras, en vez de utilizar los baremos correspondientes a la población general o a grupos particulares que se confeccionaron a partir de puntuaciones recogidas en situaciones en las cuáles las personas no tenían ningún interés en distorsionar dichas puntuaciones, esta estrategia sugiere que se confeccionen baremos a partir de las puntuaciones obtenidas en situaciones en las que las personas tengan un interés directo en distorsionar (pe., en situaciones de selección de personal). Es obvio que los baremos confeccionados de esta manera incluyen ya una parte de la puntuación normativa que corresponde a la distorsión y que es común a todos los evaluados. Por tanto, la deseabilidad habrá sido ya parcialmente neutralizada. Es decir, se habrá desarrollado y aplicado un marco común a todos los evaluados, en vez de utilizar un marco que sólo es común a las personas que no distorsionan nunca, como sería el basado en los baremos creados con una población normativa que contestó al cuestionario en una situación de nula deseabilidad social (pe., con muestras de estudiantes no inducidos a distorsionar).

### CONCLUSIONES Y SUGERENCIAS PARA LA PRÁCTICA

La posibilidad de que las respuestas a los cuestionarios de personalidad sean distorsionadas, tanto de modo positivo como de modo negativo, es real y desde hace más de sesenta años se conoce este fenómeno. Este hecho ha preocupado y sigue preocupando a los profesionales de la psicología del trabajo y las organizaciones que tienen responsabilidades en la utilización de medidas de personalidad para sus trabajos cotidianos (pe., selección de



personal). Dicha preocupación ha conducido a dichos profesionales a optar por diferentes aproximaciones a la hora de posicionarse sobre el uso de las medidas de personalidad. Entre tales posiciones ha habido tres más frecuentes: (a) la de aquellos que a la vista de la posible distorsión de las respuestas, descalificaron la utilización de las medidas, considerando que la distorsión invalidaba la capacidad predictiva de las mismas y que no se podían tomar decisiones correctas a partir de dichos cuestionarios de personalidad. Por tanto, la opción correcta era descartarlas de la bolsa de herramientas del psicólogo organizacional; (b) la de aquellos que, aun reconociendo que existía la posibilidad de distorsión, consideraban que las medidas de personalidad seguían siendo válidas y útiles para el trabajo profesional y, por tanto, deberían buscarse fórmulas para superar o neutralizar esta deficiencia. Este es el caso de los psicólogos que utilizaban medidas correctoras de las puntuaciones. (c) la última posición vendría representada por aquellos que consideraban que la deseabilidad social no era un problema importante y que sería mucho peor no utilizar medidas de personalidad en su práctica profesional.

La investigación llevada a cabo en las dos últimas décadas y especialmente en los últimos diez años, ha permitido llegar a conclusiones sólidas sobre los efectos y los posibles paliativos de la deseabilidad social. Lo siguiente es un resumen de dichas conclusiones:

- 1) Las medidas de personalidad, tanto las basadas en el modelo de los cinco grandes factores, como los cuestionarios de personalidad ocupacional orientados al criterio, como otras medidas de personalidad basadas en modelos alternativos, son excelentes predictores de desempeño ocupacional, del éxito en el aprovechamiento de la formación, de las conductas contraproductivas, del liderazgo, de la satisfacción laboral, del compromiso organizacional, de la adquisición del conocimiento y de muchos otros criterios relevantes para las organizaciones. En algunos casos, las medidas de personalidad son los mejores predictores de dichos criterios (pe., conductas contraproductivas, satisfacción).
- 2) La deseabilidad social afecta a todos los métodos de evaluación de la personalidad basados en cuestionarios, incluso a aquellos que pretendidamente no les afecta, como son los cuestionarios de elección forzosa o medidas ipsativas. Es decir, no existe nin-

guna medida auto-informe de personalidad que esta libre de deseabilidad social.

- 3) La deseabilidad social, como distorsión relevante, sólo afecta a un porcentaje pequeño de las personas evaluadas en los procesos organizacionales.
- 4) El efecto de la deseabilidad social varía según la forma elegida para suscitarse dicha deseabilidad. En las situaciones habituales de selección, la deseabilidad tiene un efecto promedio de 0.38 unidades de desviación. Puesto en puntuaciones T, hace que se incrementen o decrezcan las puntuaciones en 3.8 puntos.
- 5) La deseabilidad social se compone de dos factores, el manejo de impresiones y el auto-engaño. En los contextos organizacionales, el primero es el más relevante y el que infla (o desinfla) las puntuaciones en los cuestionarios de personalidad.
- 6) La deseabilidad social no está relacionada con el desempeño laboral, por tanto no tiene efectos sobre la validez de las medidas de personalidad. En otras palabras, si se eliminan los efectos de la deseabilidad social de las medidas de personalidad, éstas no tienen una validez más alta para predecir el desempeño ocupacional. Por tanto, la creencia mencionada más arriba, referida a que la deseabilidad social invalida la capacidad predictiva de las medidas de personalidad es simplemente errónea.
- 7) A lo largo de los años se han desarrollado diversas estrategias para neutralizar o reducir los efectos de la deseabilidad y la mayoría de ellas ha resultado ser ineficaz o contraindicada:
  - a) La estrategia de los ítems de elección forzosa no impide que sean distorsionados, no mejora la validez predictiva de las medidas, afecta negativamente a su fiabilidad, tiene serios problemas psicométricos y no permite comparar a los evaluados entre sí, por lo cual no puede ser utilizada apropiadamente en aquellas situaciones que requieran comparar a los candidatos (pe selección, formación). Por tanto, no debería ser utilizada.
  - b) La estrategia de descartar a los candidatos que puntúan alto en las escalas de deseabilidad social tiene serias complicaciones teóricas, prácticas y legales. Por tanto, no debería ser utilizada.
  - c) La estrategia de "corregir" las puntuaciones de modo subjetivo tiene complicaciones teóricas y

prácticas. Afecta a la validez de constructo y predictiva de las pruebas y es inviable con números elevados de candidatos. Por tanto, no debería ser utilizada.

- d) La estrategia de "corregir" mecánicamente (pe., mediante una ecuación de regresión) las puntuaciones adolece de las mismas complicaciones teóricas que la estrategia anterior. Por tanto, no debería ser utilizada.
  - e) La estrategia de informar a los evaluados sobre la existencia de detectores de distorsión y alertarlos sobre las posibles consecuencias de dicha distorsión reduce considerablemente la misma y es un medio barato de control de la deseabilidad social. Debería ser utilizada en todos los procesos y evaluaciones en los que la deseabilidad social pudiera estar presente.
  - f) La estrategia de crear baremos a partir de muestras de solicitantes de empleo neutraliza los efectos de la deseabilidad social. Por tanto, debería ser utilizada en todos los procesos de selección. Y los profesionales deberían utilizar en tales casos instrumentos que hayan elaborado tales baremos.
- 8) La combinación de las estrategias de informar a los evaluados y usar baremos elaborados con muestras de solicitantes produce los resultados óptimos para la reducción de la deseabilidad social. Esta última es la mejor opción para los profesionales.

Las anteriores conclusiones están basadas en la evidencia disponible en la actualidad, fundamentalmente lograda a partir de meta-análisis y de estudios con muestras grandes ( $N > 1200$ ). A la vista de las mismas, los profesionales de la psicología del trabajo y las organizaciones pueden utilizar con confianza las medidas de personalidad para tomar decisiones, con las limitaciones inherentes a cualquier medida psicológica (es decir, fiabilidad, validez y utilidad), sin que la deseabilidad social sea un problema que inhabilite su uso. Más bien parece que la cuestión de la deseabilidad social ha sido la manifestación de la creación artificial de un problema de escasa relevancia para la profesión.

## REFERENCIAS

Baron, H. (1996). Strengths and limitation of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56.

- Barrick, M.R. y Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M.R., Mount, M.K. y Judge, T. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, 69, 25-39.
- Block, J. (1965). *The challenger of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century Crofts.
- Cattell, R.B., Eber, D. y Tatsuoaka, M. (1970). *Handbook of 16PF*. Champaign, IL: IPAT.
- Christiansen, N.D., Burns, G.N. y Montgomery, G.E. (2005). Reconsidering forced-choice itemformats for applicant personality assessment. *Human Performance*, 18, 267-307.
- Christiansen, N.D., Goffin, R.D., Johnston, N.G. y Rothstein, M.G. (1994). Correcting the Sixteen Personality Factors test for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47, 847-860.
- Clarke, S. y Robertson, I.T. (2005). A meta-analytic review of the Big Five personality factors and accidents involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology*, 78, 355-366.
- Closs, S.J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, 69, 41-47.
- Cornwell, J.M. y Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Wilson (1991). *Journal of Occupational and Organizational Psychology*, 67, 89-100.
- Costa, P. T. y McCrae, R. R. (1992). *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Crowne, D.P. y Marlowe, D. (1964). *The approval motive*. New York : Wiley.
- Connolly, J.J. y Viswesvaran, C. (2000). The role of affectivity in job satisfaction: A meta-analysis. *Personality and Individual Differences*, 28, 115-126.

- lity and Individual Differences, 29, 265-281
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York : Dymden.
- Ellington, J., Sackett, P.R. y Hough, L.M. (1999). Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.
- Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin, 43*, 385-440.
- Eysenck, H.J. y Eysenck, S.B. (1964). *The Eysenck Personality Inventory manual*. London: Routledge.
- Goffin, R.D. y Christiansen, N.D. (2003). Correcting personality test for faking: A review of popular personality tests and initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340-344.
- Gough, H.G. (1987). *Manual: The California Personality Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.
- Hogan, J. y Holland, B. (2003). Using theory to evaluate personality and job performance relations: A socio-analytic perspective. *Journal of Applied Psychology, 88*, 100-112.
- Hough, L.M. (1992). The "Big Five" personality variable-construct confusion: Description versus prediction. *Human Performance, 5*, 139-155.
- Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*, 209-244.
- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. y McKloy, R.A. (1990). Criterion-related validities of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
- Hough, L.M. y Paullin, C. (1994). Construc-oriented scale construction: The rational approach. En G.S. Stokes, M.D. Munford y W.A. Owens (Eds). *The biodata handbook: Theory, research and application*. Palo Alto, CA; Consulting Psychologists Press.
- Hurtz, G.M. y Donovan, J.J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869-879.
- Jackson, D.N. (1994). *Jackson Personality Inventory*. - Revised Manual. Port Huron, MI: Sigma Assessment Systems.
- Judge, T., Bono, J.E., Ilies, R. y Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.
- Judge, T. A., Heller, D. y Mount, M. K. (2002). Five-Factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*, 530-541.
- McCrae, R.R. y Costa, P.T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.
- McFarland, L.A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality tests scores. *International Journal of Selection and Assessment, 11*, 265-276.
- Meade, A.W. (2004) Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*, 531-552.
- Meehl, P.E. y Hathaway, S.R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30*, 525-564.
- Mount, M. K. y Barrick, M. R. (1995). The Big Five personality dimensions: implications for research and practice in human resources management. In K.M. Rowland y G. Ferris (Eds.), *Research in personnel and human resources management, vol. 13* (pp 153-200). Greenwich, CT: JAI Press.
- Mueller-Hanson, R., Heggstad, E.D. y Thornton, G.C. (2003). Faking And selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.
- Ones, D.S. y Viswesvaran, C. (1998a). Integrity testing in organizations. In R.W. Griffin, A. O'Leary-Kelly y J. Collins (eds). *Dysfunctional behaviour in organizations: Vol. 2. Nonviolent behaviours in organizations* (pp. 243-276). Greenwich, CT: JAI Press.
- Ones, D.S. y Viswesvaran, C. (1998b). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269.
- Ones, D.S. y Viswesvaran, C. (2001a). Personality at Work: criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts y

- R.T. Hogan (Eds), *Applied personality psychology* (pp. 63-92). Washington, DC: American Psychological Association.
- Ones, D.S. y Viswesvaran, C. (2001b). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31-39.
- Ones, D.S., Viswesvaran, C. y Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Ones, D.S., Viswesvaran, C. y Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology (Monograph)*, 78, 679-703.
- Ones, D.S., Viswesvaran, C. y Schmidt, F.L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality*, 17, 19-38.
- Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. En H.I. Braun y D.N. Jackson (Eds.). *Role of constructs in psychological and educational measurement* (pp 49-69). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rolland, J.P. y Mogenet, J.L. (2001). *Système de description en cinq dimensions (D5D). Manuel réservé aux psychologues*. Paris: Les Editions du Centre de Psychologie Appliquée.
- Rosse, J.G., Stecher, M.D., Miller, J.L. y Levin, R.A. (1998). The impact response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Rothman, S., Meining, D. y Barrick, M.R. (2002). *Predicting performance using personality measures in South Africa*. Symposium on Personality at work in a cross-cultural context (D.S. Ones, Chair), Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.
- Salgado, J.F. (1996) Análisis exploratorio y confirmatorio del inventario de personalidad de cinco factores (IP/5F). *Psicológica*, 17, 353-366.
- Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J.F. (1998). The Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, 11, 271-288.
- Salgado, J.F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117-125.
- Salgado, J.F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323-346.
- Salgado, J.F. (2004). *Moderator Effects of Job Complexity on the Big Five Validity*. Poster presentado en la Conferencia de la Society for Industrial and Organizational Psychology, Chicago, Illinois.
- Salgado, J.F. y De Fruyt, F. (2005). Personality in personnel selection. En A. Evers, O. Schmit-Voskuyl y N. Anderson (Eds). *Handbook of personnel selection*. Oxford, UK: Blackwell.
- Salgado, J.F., Remeseiro, C. y Iglesias, M. (1996). Personality and Test-Taking Motivation. *Psicothema*, 8, 553-562.
- Salgado, J.F. y Moscoso (2000). Autoeficacia y criterios organizacionales de desempeño. *Apuntes de Psicología*, 18, 179-191.
- Seisdedos, N. (1988). Selección, cuestionarios y distorsión motivacional (actitud inteligente de adaptación). *Revista de Psicología del Trabajo y las Organizaciones*, 4, 3-33.
- SHL (1999). *OPQ32 Manual and User's Guide*. Surrey: SHL.
- Van Rooy, D. L. y Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, 65, 71-95.
- Viswesvaran, C., Ones, D.S. y Hough, L.M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings?. *International Journal of Selection and Assessment*, 9, 277-289.
- Yoo, T.Y. y Ming, B.M. (2002). *A meta-analysis of the Big Five and performance in Korea*. Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.



## EVALUACIÓN DE LAS DISTORSIONES DE RESPUESTA MEDIANTE EL MMPI-2

### ASSESSMENT OF SELF-REPORT RESPONSE DISTORSION BY MEANS OF THE MMPI-2

Héctor González Ordi e Iciar Iruarrizaga Díez

Universidad Complutense de Madrid

*El Inventario Multifásico de Personalidad de Minnesota 2, MMPI-2, es uno de los instrumentos multiescalares de amplio espectro mejor validados para explorar estilos de respuesta aleatorios, simulación de psicopatología y defensividad. En el presente artículo se exponen en detalle las distintas posibilidades, estrategias y escalas que proporciona el empleo del MMPI-2 para la evaluación de estilos de respuesta sobredimensionados (escalas F, Fb, F(p), índice F-K positivo, FBS y DsR), estilos de respuesta infradimensionados (escalas L, K, índice F-K negativo, S y Wsd) y estilos de respuesta aleatorios (escalas ?, F, Fb, VRIN y TRIN).*

*The Minnesota Multiphasic Personality Inventory 2, MMPI-2 is one of the best validated multiscale measures for irrelevant responding, malingered psychopathology, and defensiveness. The present paper presents in detail different possibilities, strategies and scales provided by the MMPI-2 for the detection of overreporting responding (F, Fb & F(p) scales, F-K index: positive, and FBS & DsR scales), underreporting responding (L & K scales, F-K index: negative, and S & Wsd scales), and random responding (cannot-say, F, Fb, VRIN & TRIN scales).*

**L**a mentira, el engaño, o la ocultación, la deformación y tergiversación de la información son conductas consustanciales al ser humano (Martínez Selva, 2005), que persiguen diversos fines en el proceso de interacción social como ofrecer una impresión positiva a los demás, auto-beneficiarse y beneficiar a otros, evitar un castigo potencial o simplemente mantener una buena interacción social procurando no herir innecesariamente los sentimientos de los demás (Vrij, 2001).

Imagínese por un momento el lector que fuéramos invadidos por seres extraterrestres que tuvieran el don de la ubicuidad y lo vieran todo, lo oyeran todo, dijeran siempre la verdad sobre todo, y se dedicaran simplemente a observarnos e inmiscuirse permanentemente en las conversaciones humanas; seguramente sucumbiríamos rápidamente en un caos absoluto al no podernos enfrentar a semejante "juego de la verdad" (por cierto, este ejemplo es el argumento central de una deliciosa novela de Ciencia Ficción escrita por Fredric Brown, "Marciano, vete a casa").

El empleo de diversas estrategias para distorsionar la información buscando un fin determinado es una parte

constituyente del proceso de interacción social, conocido, aceptado y consensuado por las partes integrantes, siempre y cuando dichas estrategias se encuentren dentro de los dominios de lo socialmente admisible (Kashy y DePaulo, 1996).

El ámbito de la evaluación psicológica mediante autoinforme no es ajeno a la existencia de distorsiones de la información producidas por diversos motivos por los sujetos evaluados: las distorsiones de respuesta (Miguel-Tobal, 1993; Baer, Rinaldo y Berry, 2003). El estudio de las distorsiones de respuesta y de las estrategias o instrumentos más eficaces para detectarlas es un campo en pleno auge y evolución que, sin duda, tiene y tendrá importantes consecuencias en la práctica clínica, forense y médico-legal.

Entre los distintos tipos de distorsión de respuestas podemos encontrar los siguientes (Baer, Rinaldo y Berry, 2003):

1. Patrones de "mala imagen", estilos de respuesta sobredimensionados, o simulación (*faking bad* o *malingering*), cuando el sujeto intenta deliberadamente crear la impresión de tener alguna alteración o deterioro mediante la exageración o fabricación de síntomas, problemas y maximizando las características negativas de si mismo.

---

Correspondencia: Dr. Héctor González Ordi. Departamento de Psicología Básica II (Procesos Cognitivos). Facultad de Psicología. Universidad Complutense de Madrid. Campus de Somosaguas s/n. 28223 Madrid. España. E-mail: hectorgo@psi.ucm.es



2. Patrones de "buena imagen", estilos de respuesta infra-dimensionados, defensividad, deseabilidad social o disimulación (*faking good*), cuando el sujeto pretende deliberadamente presentar una impresión favorable de si mismo, soslayando, negando o encubriendo síntomas, problemas y destacando sus características positivas.
3. Estilo de respuesta aleatorio, cuando el sujeto responde independientemente del contenido de los ítems, debido a dificultades de lectura y/o comprensión de los mismos, falta de cooperación, descuido, falta de concentración o presencia de estados confusionales. Dentro de este apartado, la "aquiescencia" y la "no-aquiescencia", reflejan la tendencia a responder indiscriminadamente "verdadero" o "falso", respectivamente, sin considerar el contenido de los ítems a los que se está respondiendo.

### INVENTARIO MULTIFÁSICO DE PERSONALIDAD DE MINNESOTA: MMPI, MMPI-2 Y MMPI-A

El Inventario Multifásico de Personalidad de Minnesota (MMPI), desarrollado originalmente por Hathaway y McKinley (1940) y su posterior revisión y restandarización para adultos, el **MMPI-2** (Butcher, Dahlstrom, Graham, Tellegen y Kaemmer, 1989) y adolescentes, el **MMPI-A** (Butcher, Williams, Graham, Archer, Tellegen, Ben-Porath y Kaemmer, 1992), realizada a finales de la década de los 80, es uno de los cuestionarios más ampliamente utilizados para evaluar alteraciones psicopatológicas en el ámbito clínico en general (Lubin, Larsen y Matarazzo, 1984; Piotrowski, 1998) y en el contexto forense en particular (Bartol y Bartol, 2004; Boccaccini y Brodsky, 1999).

Como veremos, el MMPI-2 incluye diversos indicadores de validez que han demostrado su utilidad en la detección de simulación (Elhai, Naifeh, Zucker, Gold, Deitsch y Frueh, 2004; Guriel y Fremouw, 2003; Rogers, Sewell, Martin y Vitacco, 2003). De hecho, según Rogers (1997), el MMPI y MMPI-2 son los instrumentos multiescalares de amplio espectro mejor validados para explorar estilos de respuesta aleatorios, simulación de psicopatología y defensividad.

Butcher y Ben-Porath (2004) enumeran algunas de las características que contribuyen a la popularidad y amplia aplicación de este instrumento de evaluación psicopatológica de amplio espectro a lo largo de más de

sesenta años de existencia: (1) incluye una amplia serie de factores psicopatológicos y de personalidad que han demostrado ser fiables, válidos y estables a lo largo del tiempo; (2) se han ido incorporando nuevas escalas para dar cuenta de los avances conceptuales en psicopatología, renovando y re-actualizando el instrumento; (3) permite contrastar perfiles individuales con una amplia base de datos normativa acumulada durante decenios de investigación; (4) permite una interpretación objetiva siguiendo normas estandarizadas; y (5) se ha traducido y adaptado a diversos idiomas y países, lo que permite la comparación cross-cultural.

La adaptación española del MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen y Kaemmer, 1999) incluye 7 escalas de validez, las 10 escalas clínicas básicas originales del MMPI con sus 31 subescalas específicas, 15 escalas de contenido y 15 escalas suplementarias de diversos autores, que se han ido añadiendo a la prueba a lo largo de los años. En total, 78 escalas y subescalas, que confieren a esta prueba una riqueza de información realmente única, amplia y diversa, tal y como puede verse en la Tabla 1.

### PROCEDIMIENTO DE EVALUACIÓN DE LAS DISTORSIONES DE RESPUESTA MEDIANTE EL MMPI-2

El protocolo de valoración de las distorsiones de respuesta que seguiremos en este artículo está basado en los pasos de evaluación de la validez del MMPI-2 propuestos por Greene (1997), que en esencia consta de cinco fases, tal y como puede verse en la Tabla 2: una vez que el MMPI-2 ha sido administrado y cumplimentado, se procede a la detección del número de omisiones (ítems no respondidos) e ítems mal respondidos por parte del sujeto; tras lo cual se valora el grado de consistencia y fiabilidad de las respuestas emitidas y, siempre y cuando las distorsiones encontradas no aconsejen lo contrario, se procede finalmente a la interpretación clínica de las escalas básicas y sus subescalas, escalas de contenido y escalas suplementarias. A continuación, nos detendremos con mayor detalle en cada una de estas cinco fases.

#### Administración del MMPI-2

El MMPI-2 en su versión completa o estándar (la más habitual y recomendable), es un cuestionario de 567 ítems, con un formato de respuesta dicotómico verdadero-falso, que está diseñado para ser aplicado en adultos ( $\geq 18$

<b>TABLA 1: MMPI-2</b>			
<b>Escalas y subescalas contenidas en la adaptación española (modificado de González Ordi y Gómez Segura, 2002)</b>			
<b>ESCALAS DE VALIDEZ</b>			
?= Interrogantes	L= Mentiras	F= Escala de Infrecuencia (Exageración de síntomas)	
K= Corrección (Defensividad sutil)	F(b)= F posterior	VRIN= Inconsistencia a respuestas variables	
TRIN= Inconsistencia a respuestas verdadero			
<b>ESCALAS CLÍNICAS BÁSICAS</b>		<b>Subescalas de Harris y Lingoes</b>	
1. Hs= Hipocondría			
2. D= Depresión		D1= Depresión subjetiva D2= Retardo psicomotor D3= Disfunción física D4= Enlentecimiento mental D5= Rumiación	
3. Hy= Histeria		Hy1= Negación de ansiedad social Hy2= Necesidad de afecto Hy3= Lasiitud-malestar Hy4= Quejas somáticas Hy5= Inhibición de la agresión	
4. Pd= Desviación Psicopática		Pd1= Discordia familiar Pd2= Problemas con la autoridad Pd3= Frialdad social Pd4= Alienación social Pd5= Autoalienación	
5. Mf= Masculinidad-Feminidad			
6. Pa= Paranoia		Pa1= Ideas persecutorias Pa2= Hipersensibilidad Pa3= Ingenuidad	
7. Pt= Psicastenia			
8. Sc= Esquizofrenia		Sc1= Alienación social Sc2= Alienación emocional Sc3= Ausencia control del yo, cognitivo Sc4= Ausencia control del yo, conativo Sc5= Aus. control del yo, inhibición defectuosa Sc6= Experiencias sensoriales extrañas	
9. Ma= Hipomanía		Ma1= Amoralidad Ma2= Aceleración psicomotriz Ma3= Imperturbabilidad Ma4= Autoenvanecimiento	
0. Si= Introversión Social		<b>Subescalas "Si"</b>	
		Si1= Timidez/autocohibición Si2= Evitación social Si3= Alienación respecto a si mismo y a los otros	
<b>ESCALAS DE CONTENIDO</b>			
ANX= Ansiedad	ASP= Conductas antisociales	FRS= Miedos	TPA= Comportamiento Tipo A
OBS= Obsesividad	LSE= Baja autoestima	DEP= Depresión	SOD= Malestar social
HEA= Preocupación por la salud	FAM= Problemas familiares	BIZ= Pensamiento extravagante	WRK= Interferencia laboral
ANG= Hostilidad	TRT= Indicadores negativos de tratamiento		CYN= Cinismo
<b>ESCALAS SUPLEMENTARIAS</b>			
A= Ansiedad	R= Represión	Es= Fuerza del Yo	
MAC-R= Escala revisada de alcoholismo de MacAndrew	O-H= Hostilidad excesivamente controlada	Do= Dominancia	
Re= Responsabilidad social	Mt= Inadaptación universitaria	GM= Rol de género masculino	
GF= Rol de género femenino	PK= Escala de estrés postraumático	PS= Escala de estrés postraumático	
MDS= Malestar matrimonial	APS= Escala de potencial de adicción	AAS= Escala de reconocimiento de adicción	

años), con un tiempo de administración estimado entre 1 y 2 horas para la mayoría de los casos. En pacientes con psicopatología severa puede conllevar un tiempo de administración más prolongado, entre 3 y 4 horas. Ex-

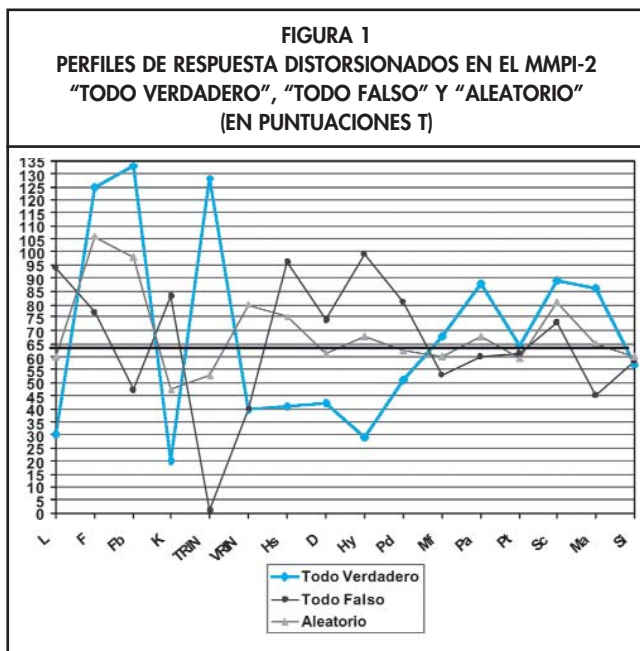
cepcionalmente existe una forma de aplicación abreviada, administrando solamente los ítems del 1 al 370, aunque resulta poco recomendable debido a que únicamente se podrían obtener resultados fiables para las escalas clínicas básicas y las escalas de validez, pero no para el resto de las escalas (Nichols, 2001).

**TABLA 2**  
**PROCEDIMIENTO DE EVALUACIÓN DE LAS DISTORSIONES DE RESPUESTA MEDIANTE EL MMPI-2 (BASADO EN GREENE, 1997)**

FASES	OBJETIVOS	ESCALAS
Fase 1	Administración del MMPI-2	Estándar: 567 ítems Abreviada: 370 ítems
Fase 2	Detección de omisiones e ítems mal respondidos	Escala de interrogantes (?)
Fase 3	Valoración de la consistencia de las respuestas	Perfiles aleatorios, Escalas VRIN y TRIN, Escalas F y Fb, tendencia a contestar verdadero o falso
Fase 4	Valoración de la fiabilidad de las respuestas: 1. Patrones de respuesta sobredimensionados 2. Patrones de respuesta infradimensionados	Escalas F, Fb, Índice F-K  Escalas L, K, Índice F-K
Fase 5	Interpretación clínica del MMPI-2	Escalas clínicas básicas y sus subescalas, escalas de contenido y escalas suplementarias

**Detección de omisiones o respuestas erróneas**

Una vez que el MMPI-2 ha sido adecuadamente cumplimentado, el primer paso en la valoración de las distorsiones de respuesta es detectar el número de omisiones o respuestas erróneas que el sujeto evaluado haya podido cometer, a través de la Escala de Interrogantes (?). Al ser una prueba larga, es frecuente que la mayoría de los individuos, presenten psicopatología o no, dejen de responder a algunos ítems o marquen ambas respuestas, verdadero y falso, erróneamente. De hecho, Greene (1997) ha estimado que el rango esperable de omisiones suele encontrarse entre 1-15 para sujetos normales y 0-20 para pacientes psicopatológicos. En general, se considera que el protocolo de administración quedaría invalidado si el individuo deja sin contestar 30 ó más ítems dentro de los primeros 370; si estas omisiones se producen a partir del ítem 370, pueden interpretarse las escalas clínicas básicas y las escalas de validez, pero no el resto de las escalas. Se considera que la omisión excesiva de ítems suele estar relacionada con patrones de defensividad, indecisión, descuido, fatiga o incapacidad para leer y comprender el contenido de los ítems (Butcher y Williams, 1992; Graham, 1993).



**Valoración de la consistencia de las respuestas**

Una vez que hemos confirmado que el número de omisiones e ítems erróneamente respondidos se encuentran dentro de los límites aceptables asegurar la validez del protocolo, la siguiente fase en la valoración de las distorsiones de respuesta en el MMPI-2 es estudiar si el sujeto evaluado ha respondido consistentemente a los ítems de la prueba. El sujeto puede responder inconsistentemente a los ítems de varias maneras: contestando preferentemente "verdadero" (aquiescencia), contestando preferentemente "falso" (no aquiescencia) o simplemente respondiendo de forma aleatoria. Los perfiles distorsionados que se obtienen en el MMPI-2 cuando se producen estas tres formas de respuesta inconsistente pueden verse en la Figura 1.

Una de las escalas más sensibles del MMPI-2 para detectar patrones de respuesta inconsistente es la Escala de Infrecuencia o escala F (Clark, Girona y Young, 2003; Sewell y Rogers, 1994), que como puede apreciarse en la Figura 1, aparece inusualmente elevada (por encima del punto de corte normativo  $T=65$ ) para las tres formas de respuesta inconsistente. La Escala F, y su homónima la Escala F posterior (Fb), son escalas diseñadas para detectar respuestas infrecuentes o respuestas que responderían afirmativamente menos de un 10% de la población normativa; en este sentido, las puntuaciones altas en F y Fb ( $T \geq 65$ ) indicarían una desviación significativa de los patrones normativos y preponderancia de estilos de respuesta no convencionales (Nichols, 2001).

Una vez confirmada la elevación significativa de la Escala F, resta averiguar cuál es la dirección del patrón de respuestas inconsistentes. Las Escalas TRIN y VRIN resultan de gran utilidad para discriminar las características del supuesto patrón inconsistente de respuesta.

La Escala TRIN (*True Response Consistency Scale*) está diseñada para detectar si existe un patrón distorsionado de respuestas aquiescente (tendencia a contestar verdadero) o no-aquiescente (tendencia a contestar falso). Como puede observarse en la Figura 1, el perfil de respuestas de "todo verdadero" se caracteriza por elevaciones muy marcadas en la Escala TRIN; mientras que en el perfil de respuestas "todo falso", TRIN presenta puntuaciones extremadamente bajas. La Escala VRIN (*Variable Response Consistency Scale*) sin embargo está diseñada para detectar específicamente estilos de respuesta aleatorios, no consistentes con el contenido de los ítems. Efectivamente, en la Figura 1 puede apreciarse que VRIN sólo aparece inusualmente elevada para el caso del perfil de respuesta "aleatorio", pero no en los casos de "todo verdadero" y "todo falso". TRIN nos confirma que el sujeto ha respondido descuidadamente a la prueba, sin reparar excesivamente en si su respuesta al contenido de los ítems era consistente o no.

### **Valoración de la fiabilidad de las respuestas**

La valoración de la fiabilidad de las respuestas pretende explorar la existencia o no de patrones de respuesta distorsionados que dificulten la adecuada interpretación clínica del MMPI-2. Básicamente, existen dos tipos de patrones a considerar en este apartado: (1) patrones de "mala imagen", estilos de respuesta sobredimensiona-

dos, o simulación (*faking bad* o *malingering*) y (2) patrones de "buena imagen", estilos de respuesta infradimensionados, defensividad, deseabilidad social o disimulación (*faking good*).

### **Estilos de respuesta sobredimensionados**

Las escalas de infrecuencia (F y Fb) han demostrado su utilidad para identificar eficazmente individuos que pretenden presentarse de forma poco favorable (mala imagen), simulando o sobredimensionando deliberadamente síntomas psicopatológicos (Bury y Bagby, 2002; Elhai, Naifeh, Zucker, Gold, Deitsch y Frueh, 2004; Graham, Watts y Timbrook, 1991; Nicholson, Mouton, Bagby, Buis, Peterson y Buidas, 1997; Strong, Greene y Schinka, 2000). De hecho, estas escalas contienen ítems seleccionados para detectar estilos de respuestas atípicos o inusuales, ya que el contenido de sus ítems está centrado en síntomas extravagantes o inusuales de psicopatología severa (Nichols, 2001). Como acertadamente indica Greene (1997), las puntuaciones altas en estas escalas pueden deberse a la existencia de estilos de respuesta inconsistente (tal y como vimos en el apartado anterior), a la presencia de psicopatología real severa, o bien a un patrón de simulación de respuestas, lo que en definitiva, sobredimensionará las puntuaciones de las escalas clínicas básicas del MMPI-2; mientras que las puntuaciones bajas suelen asociarse a la ausencia de psicopatología real, o bien a patrones de defensividad, infradimensionando las puntuaciones obtenidas en las escalas clínicas básicas del MMPI-2. Respecto a los patrones de "mala imagen" o sobredimensionados, Butcher (2005) recomienda considerar la existencia de simulación, o exageración deliberada de síntomas, cuando F y/o Fb muestren puntuaciones T mayores de 100 y VRIN sea menor o igual a 79.

Otro indicador interesante de simulación es el índice F-K de Gough (1950). Este índice se obtiene substrayendo la puntuación directa de la Escala K de validez respecto de la puntuación directa de la escala F de validez (F menos K). Si el índice es positivo a partir de un determinado punto de corte, el sujeto evidenciaría una tendencia a la exageración deliberada de síntomas (*fake bad* o mal simulador); si el índice es negativo a partir de un determinado punto de corte, el individuo mostraría una tendencia a la negación o encubrimiento de síntomas – defensividad (*fake good* o buen simulador).

Uno de los problemas del índice F-K es que no existe consenso entre los diversos autores en relación a los puntos de corte definitivos recomendados para clasificar eficazmente a individuos simuladores de individuos que no lo son, ya que dichos puntos de corte dependen en gran medida de las muestras utilizadas en su obtención. De hecho, la literatura científica señala puntos de corte recomendados para exageración de síntomas que oscilan entre +6 y +27 y entre -11 y -20 para el encubrimiento o negación de síntomas, siempre en muestras norteamericanas (ver Butcher y Williams, 1992; Greene, 1997; Meyers, Millis y Volkert, 2002; Nichols, 2001; Pope Butcher y Seelen, 1993). Respecto al uso del índice F-K en nuestro país, se han calculado específicamente los puntos de corte estimados para exageración y encubrimiento de síntomas tanto para el MMPI-2 (González Ordi y Gómez Segura, 2002) como para el MMPI-A (González Ordi, 2005) en base a las muestras de referencia que sirvieron para la adaptación española de ambas pruebas.

A pesar de que investigaciones recientes sugieren que no es más eficaz en la detección de simulación que la escala F por sí sola (Bury y Bagby, 2002; Butcher, 2005; Nicholson et al., 1997), este indicador es suficientemente sensible a la detección de la exageración de síntomas (de hecho se comporta mucho mejor que con la evaluación del encubrimiento o negación de síntomas según Nichols, 2001) como para que merezca la pena seguir teniéndolo en cuenta como información adicional en la detección de simulación, porque además correlaciona positiva y significativamente con instrumentos de evaluación de la simulación mediante autoinforme de última generación, como es el caso del SIMS - *Structured Inventory of Malingered Symptomatology* (Widows y Smith, 2005) y sigue teniendo un uso extendido en el ámbito de la evaluación forense como ayuda en la detección de la exageración deliberada de sintomatología psicopatológica (Ben-Porath, Graham, Hall, Hirschman y Zaragoza, 1995; González Ordi y Gancedo Rojí, 1999).

### **Estilos de respuesta infra-dimensionados**

Los índices del MMPI-2 más ampliamente utilizados para valorar patrones infra-dimensionados de respuesta son las escalas L y K (Baer y Miller, 2002).

La escala de Mentiras, escala L, consiste en quince ítems seleccionados con el fin de identificar a individuos que tratan deliberadamente de mostrar un patrón defensivo de

respuestas en el sentido de ocultar aquellos aspectos más negativos de su personalidad, especialmente si se obtienen puntuaciones T por encima de 66 (Butcher, 2005). Las puntuaciones T entre 60 y 65 reflejarían un intento del individuo por mostrar una imagen lo más favorable posible de sí mismo (disimulando problemas de ajuste personal o a la realidad), incapacidad para admitir las transgresiones morales más leves y un excesivo sentido de la virtud y la moralidad (Butcher y Williams, 1992; Graham, 1993).

La escala K fue desarrollada como una medida de defensividad y como un factor corrector de la tendencia de los individuos a negar la existencia de problemas psicopatológicos (Butcher, 2005). Como factor corrector, se aplica en diferentes valores a las escalas clínicas básicas Hs, Pd, Pt, Sc y Ma para ajustar su puntuación final. Como escala de validez independiente, cuando K presenta puntuaciones T entre 60 y 69 refleja la tendencia del individuo por mostrar una imagen favorable de sí mismo, minimizando su problemática en lo posible; cuando K presenta puntuaciones T igual o superiores a 70, es factible considerar que el individuo presente un patrón defensivo de respuestas (Butcher y Williams, 1992; Pope Butcher y Seelen, 1993).

Finalmente, también el índice F-K puede resultar útil como información adicional respecto a la tendencia del individuo a infra-dimensionar sus respuestas en el MMPI-2, tal y como ya se señaló anteriormente.

Es importante hacer notar en este punto que, mientras que las escalas diseñadas para explorar la tendencia a sobredimensionar las respuestas en el MMPI-2 (F, Fb, índice F-K positivo) han gozado de un mayor interés investigador y cuentan con un mayor respaldo empírico en cuanto a su eficacia para detectar la exageración deliberada de síntomas psicopatológicos y clasificar correctamente a los llamados "malos simuladores" de los no simuladores, las escalas diseñadas para detectar la tendencia a ofrecer una imagen favorable de uno mismo, disimulando u ocultando síntomas o problemática psicopatológica (L, K, índice F-K negativo) no presentan un apoyo empírico tan unánime y generalizado, y es necesario un mayor esfuerzo investigador en este sentido (ver Baer y Miller, 2002).

### **ESCALAS DERIVADAS DEL MMPI-2 PARA LA EVALUACIÓN DE LAS DISTORSIONES DE RESPUESTA**

Además de las escalas de validez incluidas rutinariamente en la versión castellana del MMPI-2, existen un



cierto número de escalas derivadas empíricamente del propio MMPI-2 pero que no formaban parte originariamente del mismo y que han que actualmente se utilizan como fuente de información adicional para la evaluación de patrones de distorsión de respuesta.

### **Escala F Psicopatológica [F(p)]**

La escala F Psicopatológica o F(p) (Arbisi y Ben-Porath, 1995) fue creada como una medida de validez adicional para explicar más específicamente las elevaciones encontradas en la escala F de validez del MMPI-2. De hecho, Arbisi y Ben-Porath (1995) sugieren que cuando F y F(p) presentan puntuaciones elevadas, es más probable atribuir con mayor seguridad dichas elevaciones a un patrón de simulación de respuestas que a la existencia de psicopatología real severa, especialmente si las escalas VRIN y TRIN no están elevadas significativamente. En este sentido, considerar conjuntamente las escalas F y F(p) resultaría más eficaz para discriminar entre grupos con psicopatología genuina de grupos de simuladores que la escala F por sí sola (Bury y Bagby, 2002; Rothke, Friedman, Jaffe, Greene, Wetter, Cole y Baker, 2000; Storm y Graham, 2000; Strong, Greene y Schinka, 2000).

### **Escala FBS (Fake Bad Scale)**

La escala FBS (Less-Haley, English y Glenn, 1991) se diseñó específicamente con el fin de ayudar a detectar simulación de quejas somáticas en el ámbito forense. Incluye ítems referidos a síntomas somáticos, alteración del sueño, síntomas relacionados con tensión y estrés, falta de energía o anhedonia, etc. Aunque recibió cierto volumen de investigación como posible escala para detección de simulación durante la década de los noventa, recientes estudios desaconsejan utilizarla como escala para detectar patrones de simulación, señalando más bien que se trataría de una escala que evalúa la tendencia a la expresión de sintomatología psicopatológica severa, centrada en los aspectos más somáticos, y distrés emocional (Butcher, Arbisi, Atlis y McNulty, 2003).

### **Escala de Disimulación de Gough revisada [DsR]**

La escala de Disimulación de Gough revisada (Gough, 1957) o escala DsR ha sido empleada en el ámbito forense para diferenciar entre sujetos que simulan síntomas de carácter neurótico, pacientes con síntomas

genuinos y población normal. Aunque su empleo está menos extendido que otras escalas para detectar estilos sobredimensionados de respuesta, como la F(p), sigue siendo utilizada como indicador adicional de posible simulación (Bury y Begby, 2002; Storm y Graham, 2000).

### **Escala Superlativa (S)**

La escala Superlativa o escala S (Butcher y Han, 1995) fue diseñada para detectar individuos que se presentan a sí mismos de manera superlativa, potenciando exageradamente los aspectos positivos de sí mismos. De hecho, correlaciona positiva y significativamente con la escala K de validez (Greene, 1997), ofreciendo información adicional sobre la tendencia a ofrecer una imagen favorable de uno mismo, disimulando u ocultando síntomas o problemática psicopatológica; en este sentido, Butcher (2005) sugiere que cuando la escala S presente puntuaciones típicas superiores a 70 se considere la posibilidad de existencia de un patrón defensivo de respuestas.

### **Escala de Deseabilidad Social (Wsd)**

La escala de Deseabilidad Social o escala Wsd (Wiggins, 1959) es un instrumento clásico en la historia del MMPI y se diseñó para evaluar la tendencia a presentarse de manera socialmente deseable. Es una de las escalas más tradicionalmente utilizadas para la exploración de los estilos infradimensionados de respuesta o patrones de defensividad.

A lo largo del texto, hemos intentado exponer las posibilidades que ofrece el MMPI-2 para explorar las distorsiones de respuesta y la simulación. A lo largo de sus 65 años de vida, el MMPI y su re-estandarización, el MMPI-2, han resultado ser instrumentos de evaluación de psicopatología en continuo cambio y renovación, con un interés especialmente centrado en la elaboración de estrategias mediante autoinforme para detección de distorsiones de respuesta, que han tenido y tienen importantes consecuencias y aplicaciones en los contextos clínico, forense y médico-legal.

El MMPI-2 en la actualidad ofrece múltiples posibilidades para la evaluación de los estilos de respuesta sobredimensionados (escalas F, Fb, F(p), índice F-K positivo, FBS y DsR), estilos de respuesta infra-dimensionados (escalas L, K, índice F-K negativo, S y Wsd) y estilos de respuesta aleatorios (escalas ?, F, Fb, VRIN y TRIN). El empleo de la in-

formación que se deriva de estas escalas permite asegurar la detección de las distorsiones de respuesta que se puedan producir cuando un individuo se enfrenta a la realización de esta prueba, especialmente si pretende simular; pero para el establecimiento definitivo de la existencia de simulación de respuesta, el profesional cualificado deberá tener en cuenta otras fuentes de información además de la aportada por el MMPI-2, ya que el estudio de la simulación requiere necesariamente una evaluación psicológica multimétodo/multisistema pormenorizada (González Ordi y Gancedo Rojí, 1999).

## REFERENCIAS

- Arbisi, P.A. y Ben-Porath, Y.S. (1995). On MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequent Psychopathology Scale F(p). *Psychological Assessment*, 7, 424-431.
- Baer, R.A. y Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: a meta-analytic review. *Psychological Assessment*, 14, 16-26.
- Baer, R.A., Rinaldo, J.C. y Berry, D.T.R. (2003). Self-report distortions (including faking, lying, malingering, social desirability). En R. Fernández-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment*. Vol. 2 (pp. 861-866). London: SAGE.
- Bartol, C.R. y Bartol, A.M. (2004). *Introduction to forensic psychology*. London: SAGE.
- Ben-Porath, Y.S., Graham, J.R., Hall, G., Hirschman, R. y Zaragoza, M.S. (Eds.) (1995). *Forensic applications of the MMPI-2*. London: SAGE.
- Boccaccini, M.T. y Brodsky, S. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice*, 30, 253-259.
- Bury, A.S. y Bagby, R.M. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment*, 14, 472-484.
- Butcher, J.N. (2005). Assessing MMPI-2 profile validity. En G.P. Koocher, J.C. Norcross y S.S. Hill III (Eds.), *Psychologists' desk reference*. 2<sup>nd</sup> edition (pp. 128-132). New York: Oxford University Press.
- Butcher, J.N., Arbisi, P.A., Atlis, M.M. y McNulty, J.L. (2003). The construct validity of the Less-Haley Fake Bad Scale. Does this scale measuring somatic malingering or feigned emotional distress?. *Archives of Clinical Neuropsychology*, 18, 473-485.
- Butcher, J.N. y Ben-Porath, Y.S. (2004). Use of the MMPI-2 in medico-legal evaluations: An alternative interpretation for the Senior and Douglas (2001) critique. *Australian Psychologist*, 39, 44-50.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. y Kaemmer, B. (1989). *MMPI-2: Manual of administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. y Kaemmer, B. (1999). *MMPI-2: Inventario Multifásico de Personalidad de Minnesota-2. Manual*. (Adaptación española realizada por A. Avila-Espada y F. Jiménez-Gómez). Madrid: TEA Ediciones.
- Butcher, J.N. y Han, K. (1995). Development of a MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. En J.N. Butcher y C.D. Spielberger (Eds.), *Advances in personality assessment*. Vol. 10 (pp. 25-50). Hillsdale, NJ: Erlbaum.
- Butcher, J.N. y Williams, C.L. (1992). *Essentials of MMPI-2 and MMPI-A interpretation*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J.N., Williams, C.L., Graham, J.R., Archer, R.P., Tellegen, A., Ben-Porath, Y.S. y Kaemmer, B. (1992). *MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Clark, M.E., Gironda, R.J. y Young, R.W. (2003). Detection of back random responding: effectiveness of MMPI-2 and Personality Assessment Inventory validity indices. *Psychological Assessment*, 15, 223-234.
- Elhai, J.D., Naifeh, J.A., Zucker, I.S., Gold, S.V., Deitsch, S.E. y Frueh, B.C. (2004). Discriminating malingered from genuine civilian posttraumatic stress disorder: A validation of the three MMPI-2 infrequency scales (F, Fp, and Fptsd). *Assessment*, 11, 139-144.
- González Ordi, H. (2005). *Informe interpretativo informatizado del Inventario Multifásico de Personalidad de Minnesota para adolescentes (MMPI-A): Manual*. Madrid: TEA Ediciones.
- González Ordi, H. y Gómez Segura, J. (2002). *Informe interpretativo informatizado del Inventario Multifásico de Personalidad de Minnesota-2 (MMPI-2): Manual*. Madrid: TEA Ediciones.

- González Ordi, H. y Gancedo Rojí, M. (1999). Evaluación de la simulación de respuesta: diagnóstico, técnicas y procedimientos. En A. Ruano Hernández, J.M. Muñoz Céspedes y C. Cid Rojo (Coords.), *Psicología de la rehabilitación* (pp. 405-441). Madrid: Fundación MAPFRE-Medicina.
- Gough, H.G. (1950). The F minus K dissimulation index for the MMPI. *Journal of Consulting Psychology, 14*, 408-413.
- Gough, H.G. (1957). *Manual for the California Psychological Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Graham, J.R. (1993). *MMPI-2: Assessing personality and psychopathology*. New York: Oxford University Press.
- Graham, J.R., Watts, D. y Timbrook, R.E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment, 57*, 264-277.
- Greene, R.L. (1997). Assessment of malingering and defensiveness by multiscale personality inventories. En R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-207). New York: The Guilford Press.
- Guriel, J. y Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A clinical review. *Clinical Psychology Review, 23*, 881-904.
- Hathaway, S.R. y McKinley, J.C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10*, 249-254.
- Kashy, D.A. y DePaulo, B.M. (1996). Who lies?. *Journal of Personality and Social Psychology, 70*, 1037-1051.
- Less-Haley, P.R., English, L.T. y Glenn, W.J. (1991). A fake bad scale on the MMPI-2 for personal-injury claimants. *Psychological Reports, 68*, 203-210.
- Lubin, B., Larsen, R.M. y Matarazzo, J.D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist, 39*, 451-454.
- Martínez Selva, J.M. (2005). *La psicología de la mentira*. Barcelona: Paidós.
- Meyers, J.E., Millis, S.R. y Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology, 17*, 157-169.
- Miguel-Tobal, J.J. (1993). Cuestionarios, inventarios y escalas. En F.J. Labrador, J.A. Cruzado y M. Muñoz (Eds.), *Manual de técnicas de modificación y terapia de conducta* (pp. 151-180). Madrid: Pirámide.
- Nichols, D.S. (2001). *Essentials of MMPI-2 assessment*. New York: John Wiley & Sons.
- Nicholson, R.A., Mouton, G.J., Bagby, R.M., Buis, T., Peterson, S.A. y Buigas, R.A. (1997). Utility of MMPI-2 indicators of response distortion: receiver operating characteristics analysis. *Psychological Assessment, 9*, 471-479.
- Piotrowski, C. (1998). Assessment of pain: A survey of practicing clinicians. *Perceptual and Motor Skills, 86*, 181-182.
- Pope, K.S., Butcher, J.N. y Seelen, J. (1993). *The MMPI, MMPI-2 & MMPI-A in court. A practical guide for expert witnesses and attorneys*. Washington, DC: American Psychological Association.
- Rogers, R. (1997). Current status of clinical methods. En R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 373-397). New York: The Guilford Press.
- Rogers, R., Sewell, K.W., Martin, M.A. y Vitacco, M.J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160-177.
- Rothke, S.E., Friedman, A.F., Jaffe, A.M., Greene, R.L., Wetter, M.W., Cole, P. y Baker, K. (2000). Normative data for the F(p) scale of the MMPI-2: implications for clinical and forensic assessment of malingering. *Psychological Assessment, 12*, 335-340.
- Sewell, K.W. y Rogers, R. (1994). Response consistency and the MMPI-2: Development of a simplified screening scale. *Assessment, 1*, 293-299.
- Strong, D.R., Greene, R.L. y Schinka, J.A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings. *Psychological Assessment, 12*, 166-173.
- Storm, J. y Graham, J.R. (2000). Detection of coached general malingering on the MMPI-2. *Psychological Assessment, 12*, 158-165.
- Vrij, A. (2001). *Detecting lies and deceit*. New York: John Wiley & Sons.
- Widows, M.R. y Smith, G.P. (2005). *SIMS: Structured Inventory of Malingered Symptomatology. Professional manual*. Lutz, FL: PAR: Psychological Assessment Resources.
- Wiggins, J.S. (1959). Interrelationship among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419-427.