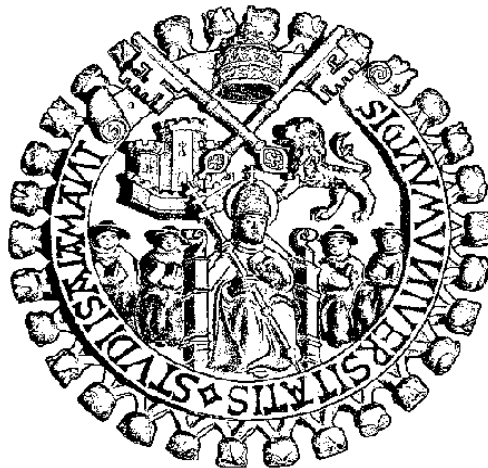


UNIVERSIDAD DE SALAMANCA

FACULTAD DE PSICOLOGÍA

Departamento de Psicología Social y Antropología



**LOS POLICÍAS COMO DETECTORES DEL
ENGAÑO: INVESTIGACIÓN EN TORNO AL
EFECTO DEL SESGO DEL INVESTIGADOR**

Hernán Alonso Dosouto

Salamanca 2009

LOS POLICÍAS COMO DETECTORES DEL ENGAÑO: INVESTIGACIÓN EN TORNO AL EFECTO DEL SESGO DEL INVESTIGADOR

**Tesis Doctoral presentada por D. Hernán Alonso Dosouto,
Licenciado en Psicología.**

D. Hernán Alonso Dosouto

A handwritten signature in blue ink, consisting of several overlapping loops and lines, positioned below the name of the author.

En Salamanca a 8 de enero del 2009

D. EUGENIO GARRIDO MARTÍN Y D. JAUME MASIP PALLEJÁ
PROFESORES DEL DEPARTAMENTO DE PSICOLOGÍA SOCIAL Y
ANTROPOLOGÍA DE LA UNIVERSIDAD DE SALAMANCA.

CERTIFICAN: Que bajo nuestra dirección, el Licenciado D. Hernán Alonso Dosouto ha realizado en el Departamento de Psicología Social y Antropología de la Universidad de Salamanca la Tesis Doctoral: "LOS POLICÍAS COMO DETECTORES DEL ENGAÑO: INVESTIGACIÓN EN TORNO AL EFECTO DEL SESGO DEL INVESTIGADOR", la cual reúne a nuestro juicio méritos suficientes de rigor y originalidad para que su autor pueda optar con ella al Grado de Doctor en Psicología.

Y para que así conste a efectos de presentación y defensa pública, firmamos la presente certificación.

En Salamanca a 8 de enero del 2009

Fdo.: Dr. D. Eugenio Garrido Martín



Fdo.: Dr. D. Jaume Masip Pallejá

ÍNDICE

	Pág.
1. Marco teórico y objetivos de esta Tesis Doctoral.....	1
EL AREA DE ESTUDIO EN QUE SE SITUA ESTA TESIS.....	1
OBETIVOS.....	2
2. Introducción.....	5
DEL ORIGEN DE LA COMUNICACIÓN Y EL LENGUAJE AL ORIGEN DE LA MENTIRA.....	5
LA MENTIRA A LO LARGO DE LA HISTORIA.....	11
LA MENTIRA EN EL MUNDO DE LAS RELIGIONES Y LOS MITOS.....	11
LA MENTIRA EN EL DESARROLLO HISTÓRICO-SOCIAL DEL HOMBRE.....	13
LA MENTIRA Y SU DETECCIÓN EN LA HISTORIA REMOTA.....	16
JUICIO MEDIANTE COMBATE O LID.....	17
LAS ORDALÍAS.....	18
La ordalía del hierro al rojo.....	19
La ordalía del agua hirviendo.....	20
La ordalía del agua fría.....	20
La ordalía del agua roja.....	20
La ordalía de mascar arroz o pan con queso.....	21
LA TORTURA.....	21
Concepto.....	21
La abolición.....	25
LA TORTURA ACTUALMENTE.....	27
3. Concepto actual de mentira.....	31
EL PROBLEMA DE LA CONCEPTUALIZACIÓN.....	32
EL TIPO DE MENTIRA CONSIDERADO EN ESTA TESIS.....	33
MENTIRAS ENTRE SERES HUMANOS.....	34
MENTIRAS DIRIGIDAS A OTRAS PERSONAS.....	34
INTENCIONALIDAD.....	35
TRANSMISIÓN DE UNA CREENCIA QUE EL MENTIROSO CONSIDERA FALSA.....	36
PREVALENCIA DE LA MENTIRA EN LAS RELACIONES HUMANAS.....	37
4. Aproximaciones actuales en la detección de la mentira...	41
TRES VÍAS PARA DETECTAR MENTIRAS.....	41
APROXIMACIÓN BASADA EN INDICADORES NO VERBALES.....	41

¿Existen razones para pensar en la existencia de indicios no verbales de engaño?.....	41
La investigación.....	42
Conclusiones.....	68
APROXIMACIÓN BASADA EN INDICADORES VERBALES.....	71
La SVA.....	72
¿Es la SVA útil para distinguir verdades de mentiras? Limitaciones.....	82
Otras herramientas para la detección del engaño desde la aproximación del contenido verbal del discurso.....	89
Conclusiones.....	112
APROXIMACIÓN BASADA EN INDICADORES PSICOFISIOLÓGICOS....	113
Cómo funciona el polígrafo.....	114
Pruebas poligráficas para la detección de la mentira.....	114
Apunte breve sobre las contramedidas.....	131
Valoración del polígrafo como instrumento para detectar mentiras según los informes de la NRC y la BPS.....	132
Métodos de detección del engaño a partir de indicadores psicofisiológicos alternativos al polígrafo.....	136

5. Los policías como detectores profesionales de la mentira y el investigador bias effect.....

LA PRECISIÓN DE LOS POLICÍAS.....	146
RESULTADOS GENERALES SOBRE PRECISIÓN.....	146
Precisión global.....	150
Precisión para las declaraciones verdaderas y falsas.....	156
Conclusiones respecto a la precisión de los policías.....	156
LA CONFIANZA EN LA PRECISIÓN DE LOS PROPIOS JUICIOS.....	159
ALGUNAS CONSIDERACIONES FINALES RESPECTO A LA PRECISIÓN.....	161
¿Existe algún grupo específico de profesionales en el que se haya observado una precisión superior?.....	162
¿Son los profesionales igualmente poco eficaces cuando juzgan la veracidad de los adultos y de los niños?.....	167
¿Influye el modo de presentación de las declaraciones en la precisión de los policías?.....	169
Conclusiones respecto a los tres interrogantes anteriores.....	170
EL INVESTIGATOR BIAS EFFECT.....	171
LA INVESTIGACIÓN DE MEISSNER Y KASSIN.....	171

6. Efecto de la experiencia profesional sobre los juicios de veracidad de los policías.....

EVIDENCIA EMPÍRICA RESPECTO A LA SOSPECHA.....	176
---	------------

SOSPECHA ESTADO, EXPERIENCIA POLICIAL, GCS, Y JUICIOS DE MENTIRA DE LOS POLICÍAS.....	185
UNA EXPLICACIÓN DE LA TENDENCIA DE LOS POLICÍAS A JUZGAR LAS DECLARACIONES COMO FALSAS BASADA EN LA GCS.....	185
Prueba empírica del planteamiento anterior.....	187
Estudio 1.....	187
Estudio 2.....	190
Estudio 3.....	195
¿POR QUÉ LOS POLICÍAS NOVELES HACEN TANTOS JUICIOS DE MENTIRA COMO LOS VETERANOS?.....	207
Estudio 4.....	209
CONCLUSIONES DE LA INVESTIGACIÓN SOBRE LA EXPERIENCIA COMO POLICÍA.....	213

7. El efecto del entrenamiento en la detección del engaño sobre los juicios de veracidad de los policías..... 216

EL ENTRENAMIENTO QUE RECIBEN LOS POLICÍAS EN LOS CENTROS DE FORMACIÓN Y DURANTE LOS PRIMEROS AÑOS COMO PROFESIONALES.....	216
LA FORMACIÓN EN DETECCIÓN DEL ENGAÑO ES ESCASA.....	217
Algunos datos sobre la formación en el Reino Unido.....	217
Algunos datos sobre la formación en Estados Unidos.....	217
Algunos datos sobre la formación en España.....	219
Conclusiones generales.....	221
EL MERCADO DE LOS MANUALES: LA <i>TÉCNICA REID</i>	222
La <i>Behavior Analysis Interview (BAI)</i> como instrumento para evaluar la credibilidad.....	223
Indicios no verbales de engaño según la <i>Técnica Reid</i>	223
El interrogatorio.....	224
¿Funciona la <i>Técnica Reid</i> ?.....	225
LOS POLICÍAS ESPAÑOLES.....	229
CONCLUSIÓN SOBRE EL ENTRENAMIENTO QUE RECIBEN LOS POLICÍAS.....	230
EL EFECTO DEL ENTRENAMIENTO EN LOS ESTUDIOS DE LABORATORIO.....	232
UNA POSIBLE EXPLICACIÓN DEL EFECTO DEL ENTRENAMIENTO: LA INFLUENCIA DE DISTINTAS MODALIDADES DE ENTRENAMIENTO SOBRE LOS JUICIOS DE VERACIDAD.....	236
ESTUDIO 5.....	237
Diseño e Hipótesis.....	237
Preparación del material.....	239
Método.....	249
Procedimiento.....	250
Resultados.....	254
Conclusiones.....	263

ESTUDIO 6.....	265
Objetivo, hipótesis y diseño.....	265
Método.....	265
Resultados.....	266
Conclusiones.....	271
CONCLUSIONES DE LA INVESTIGACIÓN REALIZADA SOBRE EL ENTRENAMIENTO EN ESTA TESIS.....	272
8. Conclusiones generales.....	276
LA EXPERIENCIA COMO POLICÍA.....	276
EL EFECTO DEL ENTRENAMIENTO.....	277
NECESIDAD DE SUPERAR EL EFECTO DEL SESGO DEL INVESTIGADOR.....	279
RESPECTO A LA EXPERIENCIA.....	279
RESPECTO AL ENTRENAMIENTO.....	280
CONCLUSIONES.....	285
BIBLIOGRAFÍA.....	287
ANEXOS.....	309

EL ÁREA DE ESTUDIO EN QUE SE SITÚA ESTA TESIS

Existe un nexo entre la Psicología y el Derecho que se hace patente en el hecho de que, tal como Garrido y Herrero (2006) señalan, ambas disciplinas tratan sobre la conducta humana. Aunque de acuerdo con Garrido y Herrero esta es una idea demasiado vaga como para constituirse en una definición de lo que es la Psicología Jurídica, también es cierto que podría ser un primer punto de partida para empezar a estudiar en qué áreas de conocimiento el derecho y psicología están, necesariamente, obligados a entenderse.

Garrido y Herrero (2006), en un intento de arrojar luz al “no siempre fácil” binomio Derecho – Psicología, ofrecen la siguiente definición de psicología jurídica:

La psicología jurídica trata de los supuestos psicológicos en que se fundamentan las leyes y quienes las aplican, bien sean juristas bien psicólogos, con el fin de explicar, predecir e intervenir (Garrido y Herrero, 2006, p. 9.)

El marco en el que se podría encuadrar esta Tesis es la Psicología Jurídica. Directamente relacionada con esta rama de conocimiento se encuentra un área de estudio que se centra en la investigación sobre testimonio. Se presenta a continuación un planteamiento teórico que enmarca específicamente a esta Tesis dentro del ámbito de estudio de la Psicología del Testimonio.

Tomaremos como base las consideraciones de Gary L. Wells. En un intento de organizar la investigación en Psicología del Testimonio, Wells (1978) diferenció entre dos tipos de variables que afectan a la calidad del testimonio: las variables a estimar y las variables del sistema. Las variables a estimar serían aquellos factores que el sistema judicial no puede controlar; en el mejor de los casos, sólo se puede determinar su presencia. Un ejemplo de variables a estimar podría ser las condiciones de iluminación o la duración del episodio

sobre el que el testigo debe testificar. Las variables del sistema son aquellos factores que sí pueden ser controlados por el sistema judicial, y cuya adecuada gestión puede facilitar que se ofrezcan testimonios de mayor calidad. Variables del sistema serían por ejemplo el tiempo transcurrido desde que el testigo presencia los hechos hasta que ha de testificar, o el tipo de interrogatorio policial utilizado para obtener una declaración.

De acuerdo con Masip y Garrido (1999) la credibilidad del testimonio, entendida como el hecho de que el testigo esté o no dispuesto a ofrecer un testimonio preciso, sería una variable a estimar, ya que depende únicamente de la voluntad del testigo, es decir, es algo que el sistema judicial no puede controlar. Dado que esta Tesis examinará a los policías como evaluadores profesionales de la credibilidad del testimonio, puede considerarse enmarcada en la Psicología del Testimonio.

OBJETIVOS

El objetivo general de esta Tesis consiste en la profundización en el estudio del *investigator bias effect* (que traduciremos como “efecto del sesgo del investigador¹”), que ha sido definido por Meissner y Kassin (2002) como la tendencia o sesgo de los policías (“profesionales” de la detección del engaño) a emitir un mayor porcentaje de juicios de mentira² cuando se enfrentan a una tarea de evaluación de la veracidad. Meissner y Kassin (2002) han planteado que en la raíz de este sesgo se encontrarían tanto el entrenamiento que los policías reciben como la experiencia policial. Es decir, lejos de mejorar la precisión de los juicios de veracidad, el entrenamiento y la experiencia no harían sino aumentar la probabilidad de que una declaración fuera juzgada como falsa. El trabajo realizado en esta Tesis se ha centrado en el estudio de cómo actúan estos dos factores para que se produzca este sesgo.

Para tratar de explicar el efecto de la experiencia policial, se parte de los estudios de Levine y McCornack (1991) sobre la Sospecha Comunicacional

¹ En este caso consideramos la palabra investigador en el ámbito de la investigación policial, y no en referencia al investigador como científico.

² A lo largo de esta Tesis se utilizará las etiquetas “juicios de mentira” para referirse a aquellas ocasiones en las que una declaración es juzgada o como falsa, y “juicios de verdad” para aquellos casos en los que se juzga que el declarante está diciendo la verdad.

Generalizada (*Generalized Communicative Suspicion* o GCS) y la Sospecha Situacional o Sospecha Estado. La GCS sería una “predisposición a creer que los mensajes generados por los demás son falsos” (Levine y McCornack, 1991, p. 328). De esta forma, se concibe como un constructo cognitivo relativamente estable a través de las situaciones. Por el contrario, la Sospecha Estado o Situacional se generaría ante determinadas claves contextuales propias de la idiosincrasia de la situación.

A partir de estos planteamientos se comprobará si la experiencia como policía, dadas las características y funciones de este trabajo, se relaciona con un aumento de la GCS, y si es así, se tratará de observar si niveles más elevados de GCS se corresponden con un mayor sesgo hacia la mentira, expresado en un mayor porcentaje de juicios de mentira ante una tarea de evaluación de la veracidad. De confirmarse estas relaciones (experiencia policial – sospecha generalizada – sesgo hacia la mentira), tendríamos una posible explicación de porqué la experiencia policial es una de las variables que está en la base del sesgo del investigador en vez de mejorar la precisión de los juicios, lo que facilitaría la búsqueda de posibles alternativas de solución.

Respecto al efecto del entrenamiento, ya he señalado anteriormente que como Meissner y Kassin (2002) han indicado, parece generar un aumento de los juicios de mentira, al tiempo que no incrementa sustancialmente la precisión global. Desde esta Tesis se investiga la posibilidad de que esto se deba al tipo de entrenamiento que se ofrece: en la inmensa mayoría de los casos se entrena a las personas para detectar la mentira. La pregunta que se constituye en el eje de la investigación realizada en este trabajo sobre el entrenamiento es: ¿qué pasaría si en lugar de enseñar a identificar claves del engaño (entrenamiento en detección de la mentira) se enseñara a identificar claves de la verdad (entrenamiento en detección de la verdad)? A partir de este interrogante se exploran los efectos de distintos tipos de entrenamiento sobre los juicios de veracidad. Encontrar que el porcentaje de juicios de verdad o mentira depende del tipo de entrenamiento recibido, sería un primer paso crucial a la hora de plantearse cómo entrenar sin sesgar los juicios.

El objetivo último de esta Tesis Doctoral es, en definitiva, tratar de contribuir desde la Psicología a un mejor ejercicio de uno de los aspectos cruciales del trabajo policial, saber distinguir cuando alguien miente y cuando dice la verdad. Para ello se intenta arrojar luz sobre dos de los más importantes factores (entrenamiento y experiencia policial) que la investigación ha revelado como condicionantes del sesgo de los policías al evaluar la veracidad de una declaración.

En este sentido, esta Tesis no supone un paréntesis o un punto de partida en la trayectoria investigadora del equipo que compone el área de trabajo en la que se enmarca. Desde el trabajo de Garrido y Masip (1999) en el que se presenta una revisión de la literatura más relevante del momento sobre los policías como detectores de mentiras, se han sucedido distintas publicaciones (véase por ejemplo Garrido, Masip y Herrero, 2004; Garrido, Masip, Herrero y Rojas, 2000; Masip, Alonso, Garrido y Antón, 2005; Masip, Garrido y Herrero, 2003c; Masip, Garrido, Herrero, Antón y Alonso, 2006) en las que se analizaron distintos aspectos de los policías como detectores de mentiras, y otras (véase por ejemplo Masip y Alonso, 2006; Masip y Garrido, 2000, 2001a,b; Masip Garrido y Herrero, 2002a,b, 2004a,b; Masip, Sporer, Garrido y Herrero, 2005) sobre distintas áreas dentro del campo de la detección del engaño. En definitiva, esta Tesis Doctoral no supone sino otro paso adelante en una línea de investigación ya sólidamente establecida dentro del Departamento de Psicología Social y Antropología de la Universidad de Salamanca.

“El hombre es por naturaleza un animal social, y el insocial por naturaleza y no por azar, o es mal hombre o más que hombre ... La sociedad es por naturaleza y anterior al individuo ... el que no puede vivir en sociedad, o no necesita nada por su propia suficiencia, no es miembro de la ciudad, sino una bestia o un dios”.

Aristóteles

DEL ORIGEN DE LA COMUNICACIÓN Y EL LENGUAJE AL ORIGEN DE LA MENTIRA

Si en algo están de acuerdo todas las disciplinas que han dedicado su estudio al ser humano, ya sea desde el prisma histórico, filosófico, religioso, comportamental, etc., es en que el hombre es por naturaleza un individuo social, y por tanto, uno de los aspectos esenciales que lo definen como Ser Humano es, precisamente, su naturaleza relacional, su inclinación natural a formar grupos con otros individuos de su misma especie. En esta dirección apuntan todos los hallazgos arqueológicos encontrados hasta el momento, lo cual no debería sorprendernos si tenemos en cuenta las evidentes ventajas adaptativas que para una especie como el hombre, o incluso los primeros protohomínidos, tenía el hecho del agrupamiento. ¿Cómo, si no es mediante la acción de equipo, se puede explicar por ejemplo la caza en una época en la que los instrumentos más modernos eran palos afilados o piedras labradas?

En este sentido, por ejemplo Beals y Hoiijer (1978) plantean la importancia del agrupamiento ya en los primeros protohomínidos, tanto como técnica de caza de mamíferos más pequeños mediante el acorralamiento, como de estrategia de búsqueda de comida y, sobre todo, como forma de protección contra animales predadores y de cuidado de las crías. Según Beals y Hoiijer (1978), sería precisamente un aumento de la cooperación social, más que el tipo de herramientas empleadas, lo que haría a estos grupos más eficaces.

Tomando pues como premisa esta naturaleza y necesidad social del hombre, podemos preguntarnos cómo o cuándo surgieron determinadas pautas de conducta social, que desde siempre han constituido la base de las relaciones humanas. Si bien parece clara, por lo menos a nivel evolutivo y adaptativo, la razón de la aparición de algunos aspectos, tales como la distribución de los roles sociales masculinos y femeninos o el establecimiento del mando, existen otros elementos indispensables en cualquier sociedad cuyo proceso de gestación resulta sin duda más complejo. Por ejemplo, ¿cómo ha evolucionado la comunicación humana desde sus orígenes hasta la formación del lenguaje tal y como hoy lo conocemos? Este es sin duda uno de los más desafiantes interrogantes, si tenemos en cuenta que la base de toda convivencia socio-relacional es la comunicación entre sus individuos. Sin embargo, actualmente no existe ninguna prueba definitiva del momento exacto en el que podemos hablar de lenguaje, entre otros aspectos, porque no está claro el momento mismo en el que se puede considerar el paso de una comunicación no verbal puramente instintiva (y en la mayoría de los casos inconsciente) al lenguaje, considerando como característica distintiva de éste la intencionalidad asociativa, que vincula la palabra o el gesto, a la realidad representada. Antes de plantear una discusión tan interesante como sin duda compleja, considero oportuno recordar que en función de los intereses de esta Tesis, y teniendo en cuenta que solamente vamos a referirnos al lenguaje humano y a las interacciones de engaño entre individuos de la especie humana, no nos adentraremos en otras disquisiciones científicas que nos alejarían de nuestro campo de estudio, como la discusión sobre comunicación, lenguaje y engaño en animales y plantas; no obstante, el lector interesado puede consultar, por ejemplo, el texto de Mitchell y Thomson (1986), quienes presentan una interesante discusión al respecto. Simplemente a modo de referencia, citaremos las posiciones de algunos autores respecto a los orígenes del lenguaje siempre en referencia al ser humano.

Según Beals y Hoijer (1978), no se dispone de ninguna evidencia directa del origen y temprano desarrollo del lenguaje, sin embargo, estos autores plantean, basándose en distintos descubrimientos arqueológicos, que los *australopithecinos* de África, formas prehomínidas, tenían pequeños encéfalos

(aproximadamente del tamaño del encéfalo de un gorila actual), que hacían muy improbable que poseyeran el lenguaje articulado, el cual, según los autores, no llegaría a existir hasta que el encéfalo del hombre se hubiese desarrollado hasta casi sus dimensiones actuales, en torno al paleolítico medio. Algunos antropólogos como Morgan (1993), sitúan el desarrollo del lenguaje a partir de formas más simples de expresión como gestos, gritos o sonidos guturales.

En relación con el desarrollo del lenguaje a partir de gestos, son interesantes los planteamientos de George Herbert Mead (1863-1931). Una de sus ideas básicas es que no puede existir una persona sin relaciones sociales, y es en el seno de estas relaciones sociales donde se plantea el concepto de “conversación de gestos”. Los gestos de un organismo se convierten en estímulos para el otro, que responde a ellos. Su respuesta se convierte en un estímulo para el primer organismo, y así sucesivamente. Según Mead, los gestos expresivos (referidos a realidades externas al individuo³) constituyen el inicio de determinadas actividades, representando por tanto dicha actividad de forma completa, aunque realmente ésta no llegue a ejecutarse hasta el final. Entre los gestos expresivos están los gestos significativos. Éstos pueden ser observados tanto por quienes los emiten como por aquellos a quienes se dirigen, y además ambos los interpretan de idéntica manera. Aparecen así los significados compartidos. Sobre esta base se construye el lenguaje humano. De hecho, las palabras son gestos significativos que permiten distinciones muy sutiles. Tales gestos también permiten el pensamiento humano, ya que Mead concibe la mente como una conversación interiorizada de gestos.

De Lumley (2000), basándose en un examen de un cráneo de un *australopithecus africanus* del que se conserva la base completa, y que data de hace unos 3 ó 3.2 millones de años, o incluso en restos de algunos *australopithecus robustus*, que son más posteriores, indica que estos cráneos carecían en su base de la flexión que permite el tracto vocal indispensable para el lenguaje articulado. Además, según De Lumley (2000), el estudio de las

³ Mead se basa en Darwin (1872), que plantea la importancia de los gestos como instrumentos de expresión de emociones internas, pero da un giro conceptual hacia Wundt, que consideró que dichos gestos no expresaban emociones internas, sino una realidad externa.

huellas endocraneanas permite observar que las zonas del lenguaje no están claramente individualizadas, en particular las áreas de Broca y de Wernicke. Por consiguiente, todo hace pensar que estos bípedos arborícolas no disponían de lenguaje articulado, aunque podían comunicarse por gestos o sonidos guturales.

Según De Lumley (2000), habrá que esperar hasta el *homo habilis*, que vivió hace unos 1.5 – 2.5 millones de años, para encontrar las zonas cerebrales del lenguaje (área de Broca en la circunvolución frontal izquierda y área de Wernicke en la circunvolución temporal izquierda). Aparecería también entonces un paladar bastante profundo, que permitiría a la lengua desplazarse y articular sonidos, junto con la ya mencionada flexión en la base del cráneo, configurando un tracto vocal mucho más humano. De Lumley (2000) plantea entonces que el *homo habilis* no deja de presentar todas las condiciones anatómicas necesarias para el lenguaje articulado. Más recientemente Montoya (2003) se refiere al *homo erectus* para hablar con cierta probabilidad de un código de signos lingüísticos que le permitiera comunicar sus sentimientos y pensamientos. Durante el paleolítico (aproximadamente 35.000 a. de J.C.), tanto el *hombre de neandertal* como el *hombre de cro-magnon* dan señales de que poseían un idioma comunicativo y una anatomía equiparable a la del hombre moderno (Montoya, 2003).

Aunque, como observamos, el origen del lenguaje articulado no se puede determinar con exactitud, sí se puede plantear el hecho de que antes de la aparición de éste ya existía algún tipo de comunicación entre los primeros homínidos (Montoya, 2003). Sin duda la comunicación no verbal, sobre todo aquel tipo de comunicación más primitiva y ligada al instinto, en la que probablemente no existía conciencia, pero que cumplía una importante función adaptativa, ha sido muy anterior al desarrollo de cualquier tipo de lenguaje articulado. Aspectos como el pelo erizado ante un peligro, taparse los ojos cuando se tiene miedo, bajar la cabeza como sumisión, o la tensión muscular y mostrar los dientes en una situación de desafío, son importantes elementos comunicacionales que se presentan de forma muy anterior a la aparición del lenguaje articulado; aspectos que sin duda cumplieron una importante función en las relaciones entre individuos, y que hoy día podemos seguir observando

como tales en los primates más cercanos al hombre. Incluso en los humanos, a pesar de haber desarrollado plenamente el lenguaje tanto a nivel verbal como no verbal, seguimos encontrando este tipo de elementos comunicacionales no verbales que se manifiestan de forma no intencional, y cuyo control exige un gran esfuerzo consciente en determinadas situaciones. Cabello erizado ante el miedo, sonrojo ante la vergüenza, apartar la mirada, tensión muscular, etc., son aspectos comunicativos que en determinadas ocasiones pueden aparecer de forma no intencional, y que precisamente por eso se han constituido en elementos clave, estudiados en distintos contextos, como en la detección del engaño. En cualquier caso, es plausible la idea de que, de forma paralela al desarrollo de la comunicación bajo cualquiera de sus formas, se han ido desarrollando otros procesos sociales que serían muy complicados sin un sistema de transmisión de ideas. Así pues, podemos plantear que la contribución de la comunicación fue decisiva para aspectos organizativos del grupo o sociedad humana, tales como la caza, el reparto de tareas, o la organización jerárquica. Y también para otros procesos más sutiles, pero no menos importantes, como procesos de influencia social tales como la persuasión, la coacción, o la influencia sobre un sujeto, y sobre todo, en los procesos de formación y cambio de la opinión del otro.

Teniendo en cuenta lo anterior, resulta cuanto menos sugerente pensar que, desde el momento (quizá no haya existido un momento puntual como tal) en que el hombre se dio cuenta de que la comunicación, o quizá deberíamos decir *el hábil manejo de la comunicación*, eran armas más poderosas que la lanza y el hacha para la obtención de recompensas sociales, comenzó un proceso ascendente e imparable a nivel comunicativo, que mucho más tarde culminaría sin duda en el arte al que la Grecia y Roma clásicas bautizarían como oratoria.

Uno de los aspectos del lenguaje y la comunicación que siempre ha sido más interesante es su empleo como herramienta de influencia social. En este sentido, desde la comunicación se puede ejercer influencia con el fin de formar o modificar opiniones para, de este modo, obtener determinadas recompensas sociales o materiales. Desde esta perspectiva, tenemos entonces que admitir que, en el complejo juego del manejo de la comunicación como estrategia de

influencia social, uno de los más eficaces recursos que se pueden emplear es el engaño.

En resumen, de lo anteriormente dicho se desprende que, aunque no sabemos exactamente cuándo el hombre empezó a organizarse en sociedad, (probablemente porque no haya un momento concreto) sí sabemos que el ser humano es un individuo social. A partir de esta idea, se plantea el hecho de que la comunicación es uno de los elementos que ha evolucionado y se ha sofisticado paralelamente al establecimiento de la organización grupal humana, y que su importancia ha sido tal que se ha constituido como herramienta principal de influencia social y obtención de metas o recompensas tan importantes como el poder y el reconocimiento de los demás. Finalmente, no cabe duda, y la historia así lo demuestra, de que la mentira es una de las estrategias que, a nivel comunicativo, siempre se ha utilizado para conseguir ejercer con éxito influencia en la formación o modificación de las creencias con el fin de obtener ciertos beneficios.

La mentira entre seres humanos es probablemente tan antigua como el momento en el que un hombre se dio cuenta del éxito del lenguaje y la comunicación allí donde fracasaban la lanza y el hacha. Desde entonces, no ha dejado de acompañarnos, tanto en los grandes acontecimientos de la humanidad, como en el devenir de la cotidianidad del día a día. No obstante, la reflexión anterior se refiere a un hipotético origen del engaño en interacciones entre seres humanos. Como ya se ha indicado, este es el tipo de mentira que se va a considerar en esta Tesis. Tratar de buscar un punto de partida de la existencia del engaño en sí mismo, se antoja imposible. En ese caso no se podrían pasar por alto los distintos estudios que han revelado la capacidad de otros primates para engañar (véase, por ejemplo, Anderson, Kuroshima, Kuwahata y Figita, 2001). Así mismo, habría que tener en cuenta los planteamientos de algunos autores (véase Mitchell, 1986) que consideran el mimetismo animal o vegetal como una forma de engaño. Considerando estos planteamientos, el engaño trascendería a la especie humana, y su origen se remontaría a un momento anterior.

LA MENTIRA A LO LARGO DE LA HISTORIA

A continuación, y a modo simplemente de ejemplo, presentamos cómo algunas mentiras se han consagrado en la historia, no sólo convirtiéndose en partes esenciales de determinados acontecimientos sociales de especial trascendencia, sino incluso introduciéndose en los mitos, enseñanzas y creencias religiosas realmente relevantes para el hombre.

LA MENTIRA EN EL MUNDO DE LAS RELIGIONES Y LOS MITOS

La mentira y sobre todo el posicionamiento ante ésta, ha formado parte esencial de las principales creencias y mitos que sustentan los credos de todo el mundo. Así, por remontarnos a una de las primeras mentiras míticas que forman parte del propio origen de la humanidad, representado según la Biblia en la religión Cristiana, tendremos que retroceder muchos siglos hasta encontramos en el Génesis, cuando la serpiente engaña a Eva:

La serpiente era el más astuto de todos los animales del campo que Yahveh Dios había hecho. Y dijo a la mujer: “¿cómo es que Dios os ha dicho: no comáis de ninguno de los árboles del jardín?”. Respondió la mujer a la serpiente: “podemos comer del fruto de los árboles del jardín. Mas del fruto del árbol que está en medio del jardín, ha dicho Dios: no comáis de él, ni lo toquéis, so pena de muerte”. Replicó la serpiente a la mujer: “de ninguna manera moriréis. Es que Dios sabe muy bien que el día que comiereis de él, se os abrirán los ojos y seréis como dioses, conocedores del bien y del mal”. Y como viese la mujer que el árbol era bueno para comer, apetecible a la vista y excelente para lograr sabiduría, tomó de su fruto y comió (Gen. 2: 1-6).

Como vemos, la serpiente miente a Eva cuando le augura una vida de Dioses si come del fruto del árbol, creando en su mente, de forma consciente, una idea que sabe que es falsa de lo que iba a ser la vida de Eva después de comer del fruto.

Kerr (1990), en relación también con los relatos bíblicos, hace referencia a uno de los episodios más conocidos del Nuevo Testamento, aquel en el que Pedro miente hasta tres veces, cuando niega conocer a Jesús.

Habían encendido una hoguera en medio del patio y estaban sentados alrededor; Pedro se sentó entre ellos. Una criada, al verle sentado junto a la lumbre, se le quedó mirando y dijo: “este también estaba con él”. Pero él lo negó: “¡mujer, no le conozco!”. Poco después, otro, viéndole, dijo: “tu también eres uno de ellos”. Pedro dijo: “¡hombre, no lo soy!”. Pasada como una hora, otro aseguraba: “cierto que éste también estaba con él, pues además es galileo”. Le dijo Pedro: “¡hombre, no sé de qué me hablas! (Mt. 26: 69-74).

Es evidente que fue el miedo el que hizo a Pedro mentir hasta tres veces respecto a su relación con Jesús, del cual, en otras partes de la Biblia, había afirmado que moriría por él.

Pero las alusiones a la mentira no son patrimonio exclusivo de la religión Cristiana. Otros textos de otros credos cuentan también con episodios de mentiras y engaños que forman una parte importante de su Tradición Sagrada. Grigorieff (1989/1995), refiriéndose al Corán, describe el pasaje del ángel Iblis, también llamado Satán, y como éste, expulsado de las alturas por una desobediencia directa de Dios, obtiene permiso divino para tentar y embaucar al hombre hasta el día de la resurrección. Grigorieff describe también a los djinns, que se presentan en el Corán como ángeles creados de puro fuego, algunos de los cuales son benéficos, mientras que otros son rebeldes, tentadores y engañadores. Finalmente, hay un pasaje en el Corán en el que, según algunos estudiosos del texto, se formula una especie de dodecálogo similar al decálogo mosaico, donde son enumerados los doce preceptos del creyente. El décimo precepto, similar al octavo mandamiento cristiano dice: “No engañar ni defraudar” (Grigorieff ,1989/1995).

De forma similar, desde el “Óctuple Sendero” de las enseñanzas budistas, el propio Buda indica las ocho direcciones justas, vinculadas entre sí, que el monje, apartado del mundo y por lo tanto de su hogar, debe tomar para llegar a la liberación total o Nirvana. La tercera de estas direcciones se refiere a La Palabra Justa: “abstenerse de mentir, maldecir, injuriar y hablar inútilmente” (Grigorieff, 1989/1995, p. 204). Por otra parte, en el plano moral y disciplinar existen diez prohibiciones explícitas para los monjes y cinco para los laicos; “no mentir” se presenta como una obligación de ambos (Grigorieff, 1989/1995).

LA MENTIRA EN EL DESARROLLO HISTÓRICO-SOCIAL DEL HOMBRE

Al margen de mitos y creencias, el desarrollo de la humanidad nos ha dejado ejemplos históricamente más contrastados de grandes mentiras cuyas consecuencias afectaron sin duda a sociedades y países enteros. Algunas de estas mentiras se recordarán por su carácter extremadamente grotesco y por el trasfondo de intereses ¿ocultos?, visibles sin embargo. Un buen ejemplo de este tipo de mentiras es sin duda el caso Taxil. Como cuenta Martín-Albo (2003), Leo Taxil, hábil escritor del siglo XIX, consiguió tener engañada a la Curia Clerical Católica por un periodo aproximado de doce años en relación con el origen y prácticas satánicas de la masonería de la época, confesando públicamente su engaño, harto de la pantomima, en una conferencia internacional en el edificio de la Sociedad Geográfica de París, ante cientos de asistentes, muchos de los cuales eran importantes cargos eclesiásticos (Martín-Albo, 2003, pp. 338 – 343).

Tampoco los políticos han pasado a la historia exentos de haber protagonizado algunos episodios de mentiras y engaño. Como observa Ekman (1985/1992), un hábil mentiroso fue sin duda Adolf Hitler. Siendo éste en Septiembre de 1938 canciller de Alemania, y Neville Chamberlain primer ministro de Gran Bretaña, se encontraron en un último intento de evitar una nueva guerra mundial. En este caso, Ekman describe como Hitler miente a Chamberlain en referencia a sus planes de invadir Checoslovaquia, dando su palabra de que si los checos satisfacen sus demandas, se mantendrá la paz. En realidad, todo es una gran mentira elaborada por Hitler, que ya había decidido atacar, pero solamente necesitaba ganar tiempo. La mentira tiene éxito y Chamberlain es engañado, con lo que trata de persuadir a los checos para que no movilicen su ejército mientras exista una posibilidad de negociar con Alemania. Después de su encuentro con Hitler, Chamberlain escribe a su hermana:

...pese a la dureza y crueldad que me pareció ver en su rostro, tuve la impresión de que podía confiar en ese hombre si daba su palabra de Honor (Ekman, 1985/1992, pp.15-16).

Muchos de nuestros políticos contemporáneos también han sido protagonistas de mentiras mundialmente conocidas; es el caso, por ejemplo, del presidente de los Estados Unidos Bill Clinton y sus declaraciones falsas en el caso Lewinsky. Según la revista digital QuePasa (“Clinton: La Hora de la Verdad”, 21 de septiembre de 1998), la Oficina del Fiscal Independiente (OFI) de los Estados Unidos presentó información sustancial y creíble contra Bill Clinton en relación con sus declaraciones de enero de 1998, cuando el presidente fue interrogado en el proceso Paula Jones y negó haber mantenido relaciones sexuales con la becaria de la Casa Blanca Mónica Lewinsky. La OFI acusaba a Clinton, entre otras cosas, de mentir bajo juramento ante un Gran Jurado. Una de las pruebas concluyentes presentadas contra el presidente Clinton fue el resultado del análisis de unas manchas encontradas en un vestido que la becaria llevaba puesto en uno de los “encuentros” con el presidente. Según los análisis se trataba de restos de semen. Posteriormente el médico de la Casa Blanca confirmó al FBI que la fuente de ADN de estos restos era el Presidente Clinton. Más adelante el propio presidente Clinton llegó a admitir, en su último día como jefe de estado, que premeditadamente había ofrecido declaraciones falsas, confusas y esquivas en relación con el Caso Lewinsky (Prensa Web, “Bill Clinton Admite que Mintió”, 20 de enero de 2001).

Otro tipo de mentiras es aquel en el que no solamente aparece una persona como directa y exclusivamente responsable del engaño, sino que todo un equipo de gobierno se ve implicado al respecto. Un ejemplo es el caso de la cadena de mentiras del gobierno Bush en relación con la guerra de Irak tras los ataques del 11 de septiembre sobre las Torres Gemelas en Nueva York (Estados Unidos). Según un análisis de la Agencia Digital Infotk (“Veinte Mentiras Sobre la Guerra de Irak”, 20 de octubre del 2003), los dos pilares que justificaban la invasión de Irak eran a su vez dos contundentes afirmaciones de la administración Bush: en primer lugar se acusaba a Irak de colaborar con Al Qaeda en los ataques del 11 de septiembre, y en segundo lugar, se acusaba a Saddam Hussein de poseer armas de destrucción masiva, lo que suponía según Bush una amenaza mundial. Bajo estas premisas, Estados Unidos invade Irak.

Actualmente, se sabe que tras invadir Irak el ejército aliado no ha encontrado armas de destrucción masiva, como tampoco las habían encontrado las numerosas inspecciones anteriores de la ONU (según Hans Blix, jefe de inspectores de armas de la ONU, más de 400 inspecciones, de las cuales muchas fueron por sorpresa). Respecto a la implicación de Irak en los atentados del 11 de septiembre, el argumento principal sobre el que se basaba esta acusación radicaba en una supuesta reunión en Praga entre Mohamed Atta, líder de los secuestradores aéreos de los atentados de dichos atentados, y un funcionario de inteligencia irakí. Sin embargo, posteriormente, los servicios de inteligencia checos admitieron que, con toda probabilidad, el contacto del funcionario iraquí no era Atta. A pesar de todo, la administración Bush no cesaba de publicar informes, basados en esta reunión, en los cuales se implicaba a Irak directamente en los atentados del 11 de septiembre.

Tampoco los políticos españoles se han librado de acusaciones de falsedad y engaño. Un ejemplo lo constituyen las que se lanzaron contra el entonces Ministro de Interior D. Ángel Acebes por sus declaraciones en los medios informativos nacionales sobre la autoría de los atentados terroristas de Madrid del 11-M. Según Rodríguez e Yoldi (2004), ya desde un primer momento de la mañana del 13 de marzo las investigaciones sobre la autoría de los atentados del 11-M apuntaban inequívocamente al terrorismo islámico. A pesar de todo, ese día el entonces Ministro de Interior Ángel Acebes mantenía una conferencia de prensa a las 14:30 en la que afirmaba que la prioridad en la investigación de los atentados era ETA, y que no le cabía ninguna duda al respecto. Sin embargo, en ese momento no solamente toda la investigación apuntaba al terrorismo islámico, sino que, además, la policía ya tenía cercados a los presuntos terroristas (tres marroquíes y dos indios). Recientemente, el periódico El País, en su formato digital, ha publicado un artículo en el que se hace referencia a la resolución definitiva del Tribunal Supremo sobre el caso 11-M cuyo encabezamiento reza:

El atentado del 11 de marzo de 2004 en Madrid, que causó la muerte a 191 personas, tiene ya un culpable: la célula islamista juzgada durante cuatro meses y medio por la matanza. El tribunal ha desmontado las principales teorías de la conspiración, incluida la más importante, el origen del explosivo

empleado en el ataque. Y ha sido concluyente: "Ninguna prueba avala la tesis de que ETA participó en el atentado. Toda o gran parte de la dinamita usada en el atentado y empleada por los terroristas que se suicidaron en Leganés procedía de la mina Conchita", ha asegurado el presidente del tribunal, Javier Gómez Bermúdez ("Ninguna Prueba Apunta a ETA", sin fechar).

Recapitulando todos estos ejemplos, y otros más que podríamos citar (pero no lo haremos por cuestiones de extensión), no podemos sino concluir que la mentira ha sido una de las grandes protagonistas de la historia de la humanidad. La mentira ha facilitado o dificultado las relaciones sociales, familiares, nacionales e internacionales; la mentira ha evitado o precipitado cambios y reformas a todos los niveles, ha mantenido unos credos e ideales, y ha derribado otros. En definitiva, ha sido una de las compañeras inseparables del complejo y entramado devenir del curso de muchos acontecimientos históricos, algunos de los cuales han determinado y seguirán determinando las formas de organización social que hoy conocemos.

Dada pues la importancia y las consecuencias del "bien mentir" para el que miente, y de la detección del engaño a tiempo para la potencial víctima del mismo, resulta evidente que si la historia ha contado con grandes y hábiles mentirosos también ha desarrollado, por el contrario, multitud de técnicas y "recetas" más o menos eficaces para descubrir la mentira. A continuación, se presenta un pequeño repaso tanto de estas técnicas como de los "expertos" en su aplicación a lo largo de la historia. Siguiendo a Masip (2002), la clasificación que se presenta divide estas técnicas básicamente en dos grupos: métodos históricos de detección del engaño, y técnicas contemporáneas o pertenecientes a la historia reciente. Nos centramos a continuación en los métodos históricos.

LA MENTIRA Y SU DETECCIÓN EN LA HISTORIA REMOTA

Dentro de este apartado, nos basaremos en la clasificación de Larson (1969), que describe tres formas generales históricamente empleadas en la detección del engaño: la prueba del combate o lid, las ordalías y la tortura.

Uno de los aspectos que caracteriza a estas formas históricamente más primitivas de detección del engaño (o más concretamente, de detección de la culpabilidad) es que todas ellas tenían como punto de apoyo y justificación la creencia en una intervención de tipo sobrenatural o divino que se manifestaría en favor de la verdad y del sujeto inocente (véase Masip, 2002; Masip, Garrido y Herrero, 2003a). El proceso consistía en someter a los sospechosos a determinadas pruebas físicas, de las que saldrían indemnes si eran realmente inocentes. El carácter sagrado de la prueba hacía que los resultados de los distintos procesos fueran casi siempre inapelables. Adicionalmente, dado que eran los “ministros terrenos” de la autoridad divina quienes tenían la última palabra respecto al acusado, existía, sin duda, un elevado índice de arbitrariedad.

JUICIO MEDIANTE COMBATE O LID

Esta práctica consistía en que, cuando era difícil determinar cuál de dos partes en disputa tenía razón, éstas debían enfrentarse en un combate o duelo, fuera éste sostenido por los propios implicados o por alguien que los representara. Se asumía que el vencedor demostraría con su victoria estar en posesión de la verdad, ya que Dios solamente da la victoria a los inocentes. Larson (1969) sitúa una de las primeras alusiones a la práctica de duelo o lid en tiempos de Tito Livio, quien describe cómo algunos habitantes de la Península Ibérica aprovechaban la oportunidad de una exhibición de lucha entre gladiadores organizada por Escipión para resolver las disputas civiles mediante el combate, en aquellos casos en los que se carecía de cualquier otro método adecuado de resolución.

Larson (1969) presenta varios ejemplos de situaciones en las que se empleaba el duelo para esclarecer la verdad de los hechos. Uno de estos ejemplos es el caso Anselmo, un ladrón sacrílego que robó los cálices de una iglesia y luego los vendió a un comerciante haciéndole jurar que mantendría en secreto tal venta. Poco después, y ante amenaza de excomunión, el comerciante confesaría a las autoridades de la iglesia todo lo ocurrido. Dado que Anselmo negó tal acusación, se recurrió al duelo para esclarecer los hechos. Al finalizar la lucha, Anselmo venció y mató al comerciante, con lo que

se determinó que era inocente de la acusación de robo. Lo curioso de la situación fue que posteriormente Anselmo, sometido a la ordalía del agua hirviendo (las ordalías eran otro tipo de métodos de detección de la verdad que examinaremos en el próximo apartado) confesó haber cometido el robo. Ante esta situación, la explicación que se ofreció para preservar la validez del duelo, fue que, sin duda, el comerciante había muerto por traicionar el juramento que había hecho a Anselmo de no decir nada de la venta de los cálices robados (Larson, 1969, p. 68).

Poco a poco, sin embargo, los principales países europeos fueron aboliendo el duelo siguiendo los dictámenes del Clero Católico, que se oponía a esta práctica. De esta forma, en el Concilio de Trento se estableció que aquellos que tomaran parte en un duelo serían excomulgados de inmediato, al igual que aquellos que aconsejaran su práctica o estuvieran presentes en este tipo de acontecimientos. Además, si uno de los duelistas resultaba muerto en la lid sería privado de un funeral cristiano. El papa Benedicto XIV decretó que a los contendientes se les podría denegar el entierro cristiano incluso aunque no muriesen en el duelo (Masip et al., 2003a, p. 11). La práctica del duelo en Hungría se eliminó en el año 1492, en Italia en 1505, en Flandes a finales del siglo XVII, y en Rusia en 1649. En Inglaterra el duelo se prolongaría hasta 1818.

LAS ORDALÍAS

Las ordalías consistían, básicamente, en someter al sospechoso de engaño a una peligrosa prueba. Si salía indemne de la misma, su inocencia quedaría entonces probada; de lo contrario, sería juzgado culpable y ajusticiado. En concreto, según Tomás y Valiente (1994) las ordalías consistían en:

...invocar y en interpretar el juicio de la divinidad a través de mecanismos ritualizados y sensibles, de cuyo resultado se infería la inocencia o la culpabilidad del acusado. No cabe duda del carácter mágico e irracional de estos medios probatorios... (p. 206).

Como se ha dicho anteriormente, la explicación subyacente a las ordalías apelaba a fuerzas divinas o misteriosas, que actuarían a favor del mantenimiento de la justicia mediante diversas manifestaciones dotadas de

sacro significado para las gentes de la época o del lugar. Las ordalías eran de uso común en la Europa alto-medieval. Según Tomás y Valiente (1979) “la sociedad alto-medieval era una sociedad teocéntrica: creía vivamente en un Dios autor de todo lo existente, eje y centro de todo lo real” (p. 134). Un Dios que era Creador y que era también Juez y “como tal lo representan los artistas medievales en los pórticos de las iglesias y colegiatas románicas: Dios ejerciendo justicia, juzgando” (Tomás y Valiente, 1979, p. 135).

Trovillo (1939) sostiene que la técnica de la ordalía no se basa en ningún conocimiento de los procesos psicológicos que subyacen al hecho de saberse culpable, sino que se deriva de la superstición y la fe religiosa. En concreto, según Trovillo (1939), en los intentos medievales de detección del engaño no existía ni un método científico, ni una técnica científica ni un conocimiento psicológico básico para establecer los parámetros de las pruebas de detección.

Finalmente, en el siglo IX, S. Abogardo, obispo de Lyon, condenó las ordalías sobre la base de que eran contrarias al espíritu evangélico de humanidad, fraternidad y paz. Abogardo planteó que las ordalías eran denominadas juicios de Dios falsamente, porque Dios nunca las prescribiría o aprobaría (Schaff, 1910). En el cuarto concilio de Letrán en 1215, Inocencio III promulgó un decreto contra las ordalías. Siguiendo esta resolución, varios sínodos de los siglos XIII y XIV anunciaron que aquellos que participasen en ordalías serían excomulgados (Kirsch, 1911). No obstante, algunas ordalías, como la del agua fría, perdurarían hasta mucho más tarde (véase Masip et al., 2003a, p. 18). A continuación revisaremos algunas de las ordalías que eran más frecuentemente empleadas.

La ordalía del hierro al rojo

Ésta fue una práctica que el cristianismo medieval tomó de la India (véase Trovillo, 1939; Larson, 1969; Masip, 2002; Masip et al., 2003a). En la India, la ordalía del hierro al rojo solía tomar la siguiente forma: el acusado se bañaba y realizaba ciertas ceremonias religiosas, tras las cuales se frotaba las manos con salvado, y se le ponían siete hojas de higuera, árbol sagrado para los hindúes, sobre las palmas. Luego, éstas se envolvían siete veces con seda.

Entonces, el acusado debía asir un hierro de cierto peso al rojo vivo, y tenía que andar trazando siete círculos. Cada uno de ellos tenía que ser mayor, en determinada cuantía, que el precedente. Si al término de este ejercicio el sospechoso no había recibido quemadura alguna en sus manos, entonces se consideraba inocente. De lo contrario se le juzgaría culpable (Larson, 1969).

La ordalía del agua hirviendo

Como señalan Masip et al. (2003a), esta forma de ordalía estaba emparentada con la anterior por el hecho de que, en ambos casos, el acusado debía ser capaz de superar ilesa altísimas temperaturas que pudieran quemar su cuerpo. La creencia subyacente consistía en que Dios evitaría que el inocente se quemase al contacto con el agua hirviendo, demostrando así su inocencia.

La ordalía del agua fría

Una curiosidad de esta práctica que la diferencia de otras ordalías, es que en este caso, el milagro se esperaba que ocurriese para condenar al culpable, no para salvar al inocente, con lo que si no ocurría milagro, y los hechos seguían el curso natural, el sospechoso se declaraba inocente. El proceso consistía en coger al acusado y sujetarlo con cuerdas de forma que pendiera sobre un tanque de agua. Entonces se iba soltando cuerda, descendiendo al sospechoso sobre el agua. Si al alcanzar el nivel del agua el sospechoso flotaba, se entendía que era culpable del crimen de que se le acusaba, mientras que se si hundía quedaba demostrada su inocencia. La explicación que se ofrecía se basaba en la fe en que el agua, elemento puro, sobre todo tras su previa bendición por parte del clérigo, no acogería en su seno a quien fuera culpable de un crimen de falso juramento (Masip et al., 2003a).

La ordalía del agua roja

Según Trovillo (1939), este procedimiento proviene de África occidental. El acusado tenía que guardar ayuno durante doce horas. Posteriormente ingería una pequeña cantidad de arroz, y luego debía beber una gran cantidad de agua (a veces hasta cuatro litros y medio) enrojecida por cierta corteza. Si la

infusión actuaba como un vomitivo, haciendo que el reo expulsara todo el arroz, su inocencia quedaba demostrada. Ahora bien, si el acusado no vomitaba el arroz, o si el efecto de la infusión era el de un laxante, entonces el reo era condenado. La explicación ofrecida sobre esta prueba, era que un fetiche de la víctima entraba en el cuerpo del acusado con el agua roja, examinaba entonces el corazón del reo, y si lo encontraba culpable, expulsaba el arroz como evidencia.

La ordalía de mascar arroz o pan con queso

Esta ordalía, aunque de origen hindú, estaba bien establecida entre los representantes del clero Católico Romano hacia el año 1150. En el *Ayeen Akbery* (citado por Trovillo, 1939) se describe la forma original hindú en los siguientes términos:

La ordalía se realiza con un tipo particular de arroz denominado *shatee*, que se prepara con varios encantamientos. La persona que va a ser juzgada lo come, con el rostro hacia el este, y entonces escupe. Si la saliva está mezclada con sangre, o las esquinas de su boca se hinchan, o si el acusado tiembla, entonces es declarado mentiroso (Trovillo, 1939, p. 853; ver también Larson, 1969, p. 77).

En la Europa medieval, esta práctica se ejercía con una onza de pan, a menudo acompañada con algo de queso. La ceremonia incluía determinadas prácticas religiosas, tras las cuales el acusado debía comer el pan. La creencia era que si éste era culpable, Dios enviaría al ángel Gabriel que obstruiría su garganta de modo que no podría tragar el alimento (Véase Masip et al., 2003a, p. 20).

LA TORTURA

Concepto

Desde los juristas romanos de los siglos II y III hasta los historiadores de nuestros días, quienes más atención han dedicado a establecer una definición de la tortura se han basado en concepciones notablemente similares. Peters (1987) recoge algunas de las definiciones históricas de tortura. Así, según el jurista del siglo II Ulpiano:

Por *quaestio* (tortura), hemos de entender el tormento y sufrimiento del cuerpo para obtener la verdad. Ni el interrogatorio en sí mismo, ni el temor ligeramente inducido, se relacionan en verdad con este edicto. Por lo tanto, puesto que la *quaestio* debe ser entendida como violencia y tormento, estas son las cosas que determinan su significado (Peters, 1987, p. 12).

Tomás y Valiente (1994), refiriéndose al concepto de tortura judicial en Castilla durante los siglos XVI y XVII, plantea a modo de síntesis que:

El tormento era una prueba del proceso penal, subsidiaria y reiterable, destinada a provocar por medios violentos la confesión de culpabilidad de aquel contra quien hubiera ciertos indicios; o dirigida, a veces, a obtener la acusación del reo contra sus cómplices, o también a forzar las declaraciones de los testigos (Tomás y Valiente, 1994, p. 99).

La convención contra la tortura de las Naciones Unidas, define en su artículo 1 la tortura como:

Cualquier acto mediante el cual se infringe dolor o sufrimiento severo, sea éste físico o mental, a una persona, con propósitos tales como obtener de él o de una tercera persona información o una confesión, castigarlo por un acto que él o una tercera persona ha cometido o se sospecha que ha cometido, o intimidar o coaccionar a esta persona o a una tercera persona, o por cualquier razón basada en discriminación de cualquier tipo, cuando este dolor o sufrimiento se infringe por parte o bajo la instigación o con el consentimiento o aquiescencia de una autoridad pública u otra persona que actúe con capacidad oficial (Amnesty International, 2000).

En esta definición, a diferencia de las anteriores, ya se hace referencia de forma explícita no solamente al dolor o sufrimiento físico, sino también al mental.

Considerando la clasificación de métodos históricos de detección del engaño de Larson (1969), a primera vista parece sorprendente el hecho de que el autor incluyera a la tortura en su clasificación dentro de esta categoría (es decir, de los procedimientos históricos). Como argumentan Masip et al. (2003a), la edición original del libro de Larson data de 1932, es decir, de antes de la Segunda Guerra Mundial, momento en que los nazis emplearían la tortura ampliamente. Por otro lado, aún hoy el uso de la tortura aparece como

algo común en tiempos de guerra. Y en la época del propio Larson era utilizada con cierta frecuencia por las fuerzas de la ley. Por lo que cuenta ese autor, no era extraño que en su época los policías instruyeran a los nuevos reclutas sobre cómo infringir el mayor daño físico posible sobre un sospechoso que no quisiera hablar de manera tal que no se produjeran señales visibles tales como moratones o desgarros en la piel. Esta práctica se conocía como “tercer grado”, consistiendo por tanto en un sistema de obtención de información y elicitación de confesiones de los sospechosos a través de distintos tipos de tortura.

Con referencia al tercer grado, Larson (1969) describe uno de los ejemplos más claros de casos en que la tortura física se acompaña también de brutalidad psicológica para forzar una confesión. Se trata del caso de un jurado americano que, tras conocer los métodos empleados para obtener de un sospechoso (que inicialmente se había declarado inocente) una detallada confesión sobre cómo se había producido el asesinato de una joven, lo absolvió.

...los detectives habían llevado a un hombre joven andrajoso a las “habitaciones de atrás”. Lo situaron al lado del ataúd, súbitamente retiraron la sábana, mostrándole así el rostro pálido contusionado y brutalmente lacerado... (del cadáver de la joven). La pregunta inmediata fue: ¿Cuándo la viste? El sospechoso cayó de rodillas y se cubrió el rostro con las manos, pero los policías lo arrastraron de nuevo hasta el ataúd y le obligaron a poner su mano derecha sobre el cuerpo del cadáver. El sospechoso obedeció temblando, y al instante siguiente, se desmayó. Los detectives lo pusieron en pie de nuevo, y comenzaron a disparar pregunta tras pregunta mientras lo forzaban a acariciar el pelo y la mejilla del cadáver. El sospechoso, evidentemente sin control de su mente, confesó haber cometido todas las torturas que los policías le indicaron que había cometido con la joven, y en su estado de demencia, incluso añadió detalles a la falsa historia (Larson, 1969, p. 96).

Este ejemplo es suficientes para cuestionarnos por qué, entonces, Larson consideraba la tortura como un método histórico junto con otros procedimientos como las ordalías y el duelo. Si bien en su tiempo las ordalías y

los duelos ya no eran entonces prácticas empleadas, la tortura se seguía utilizando.

Una explicación válida de esta clasificación es que el autor, sin duda, a la hora de considerar los métodos de detección del engaño, tuvo más en cuenta el momento en el que los distintos métodos habían empezado a emplearse, y no si en su tiempo se seguían empleando. Desde esta perspectiva, hay que tener en cuenta que la tortura es, históricamente hablando, uno de los métodos más antiguos que han existido para la detección de la culpabilidad y la obtención de una confesión (véase Masip et al., 2003a, p. 40). Así por ejemplo, Según Larson (1969), ya en la Grecia clásica el uso de la tortura aparece claramente establecido. Aunque como regla general ningún hombre libre debería ser torturado, en ningún caso los esclavos entraban dentro de esta categoría.

Peters (1987) plantea cómo en la Grecia clásica uno de los elementos diferenciales más importantes en la aplicación de la justicia era la honorabilidad del acusado. Así, el honor del hombre que ostentaba la categoría de ciudadano libre daba un gran valor a su palabra jurada. Por el contrario, quienes carecían de privilegios legales (los esclavos) debían ser obligados a entrar en un estatus especial en el que su testimonio se pudiese considerar aceptable. Era por medio de la coerción física que el testimonio de los esclavos se hacía igual al de los ciudadanos libres.

En Roma, en un momento inicial, los principios fueron similares a los de Grecia, aunque posteriormente cualquier ciudadano sería considerado responsable de sus actos, y la tortura sería empleada por los emperadores en cualquier caso en el que se deseara obtener una confesión sin que se establecieran diferencias sociales. De esta forma, la ley romana admitía la tortura en aquellos casos de acusaciones de delitos contra el estado o la persona sagrada del emperador.

Como podemos ver, tanto en la Grecia clásica como en Roma aparece la tortura como una práctica establecida institucionalmente, con lo que es perfectamente justificable que Larson (1969) la incluyera dentro de lo que él

denominó métodos históricos de detección del engaño. Hay además una segunda razón que justifica el lugar que ocupa la tortura en la clasificación de Larson, y es el hecho de que la tortura, al igual que las ordalías o el duelo, es un método que no tiene ninguna base científica en la que sustentarse. No sería lícito pues situar a la tortura junto a otros métodos de detección del engaño que sí se han desarrollado partiendo de un análisis científico del problema, solamente porque su uso se haya prolongado en el tiempo, llegando a coincidir en este sentido con dichos métodos científicos. La diferencia, pues, entre la tortura y otras formas de detección citadas, como las ordalías, es que el empleo de la misma, de forma directa u oculta bajo nombres más eufemísticos, como “tercer grado”, se ha prolongado mucho más en el tiempo (véase Masip, 2002).

La abolición

Según Peters (1987), en la mayoría de estados de la Europa de finales del siglo XVIII el proceso de la abolición legislativa de la tortura fue, primero, parte de una revisión general de los sistemas jurídicos penales, y segundo, un proceso que se produjo en largos periodos de tiempo, comúnmente durante varias décadas, no instantánea ni categóricamente. Aunque algunos estudiosos han visto la causa de la abolición de la tortura en el pensamiento ilustrado de la época, otros historiadores como Langbein (1976) plantean otro tipo de razones, basándose en la observación de que muchas de las ideas sobre la tortura de filósofos ilustrados como Voltaire, Beccaria o Verri, ya habían sido postuladas muchos siglos antes. Por ejemplo, ya en el siglo II Quintiliano había aludido al peligro de generar falsas confesiones con el uso de la tortura.

Según Langbein (1976), la verdadera causa que supondría un primer paso en la abolición de la tortura sería un cuestionamiento general del sistema penal de la época, en el que se hallaba inserta la tortura. Este autor observa que se produce el paso de un sistema penal medieval centrado esencialmente en “la prueba”, que no admitía evidencias circunstanciales, y que buscaba la demostración rotunda de la culpabilidad, por ejemplo mediante la confesión del acusado, a un sistema penal nuevo centrado en “la sanción”, en el que se

ponía más énfasis en las nuevas modalidades de castigos, alternativos a la pena capital y más o menos severos dependiendo de la convicción subjetiva del jurado de la culpabilidad del acusado. Langbein (1976) plantea que este nuevo sistema, al no estar tan centrado en la prueba de culpa, no exigía ya en todos los casos una confesión del acusado, con lo que la tortura comenzaba a dejar de ser un elemento esencial.

Peters (1987) señala que fue Suecia el primer país que jurídicamente abolió la tortura. En Prusia, Federico I, en 1721, sostuvo que la tortura solamente podía ser aplicada después del consentimiento del monarca en cada caso particular. En 1754, todo proceso de tortura fue abolido en Prusia. Entre 1738 y 1789, el reino de las dos Sicilias pasó por un proceso similar, al igual que el ducado de Baden entre 1767 y 1831, y los Países Bajos Austríacos entre 1787 y 1794. Venecia suprimió la tortura entre 1787 y 1800, y Austria entre 1769 y 1776. Otros reinos y principados abolieron esta práctica como rutina entre el último cuarto del siglo XVII y primer cuarto del siglo XIX. Brunswick, Sajonia y Dinamarca lo hicieron en 1770, Mecklemburgo en 1769, Polonia en 1776, Francia en 1780, Toscana en 1786, Lombardía en 1789, Noruega en 1819, Portugal en 1826 Grecia en 1827 y los Países Bajos en 1798 (Peters, 1987). En lo que se refiere a nuestro país, Según Tomás y Valiente (1994), el primer texto legal español que abolió la tortura fue la Constitución de Bayona de 1808. En sus dos primeros proyectos no se mencionaba el tormento, pero en el tercero se reprodujo casi textualmente el artículo 82 de la Constitución francesa que hacía referencia a la abolición del tormento (pues en Francia fue la monarquía absoluta la que tomó esta medida). En este fragmento del texto de Bayona, se prohibía cualquier rigor ilegal empleado en la prisión o detección de los reos. Ésta sería la puerta por la que se introduciría finalmente la abolición de la tortura. El artículo 114 del tercer proyecto pasó a ser el 133 de la constitución, con esta redacción definitiva:

El tormento queda abolido; todo rigor o apremio que se emplee en el acto de la prisión o en la detención y ejecución, y no esté expresamente autorizado por la ley, es un delito (Tomás y Valiente, 1994, p. 136).

LA TORTURA ACTUALMENTE

A pesar de las reseñas referidas a la abolición legal de la tortura que acabamos de exponer, existen datos que indican que la situación al respecto sigue siendo preocupante. Masip (2002), a partir del análisis de distintos informes de Amnistía Internacional, señala en su tesis doctoral que a finales del siglo XX, la tortura o los malos tratos se han seguido aplicado en 150 países (véase Masip, 2002). Actualmente, en pleno siglo XXI, el panorama mundial no parece haber mejorado al respecto.

Amnistía Internacional ha publicado recientemente un informe en el que se presenta la situación que se vive actualmente en el mundo respecto a los derechos humanos (Amnistía Internacional, 2008). Exponer aquí todos los datos referidos a la tortura que se incluyen en dicho informe alargaría en exceso este apartado. Dado que el informe se organiza de forma territorial el resumen que se presenta a continuación sigue un orden similar.

- a) África. En múltiples países (por ejemplo Angola, Benín, Burundi, Eritrea, Guinea Ecuatorial, Kenia, Mauritania, Mozambique, Níger, Sudán y Zimbabue), se denuncia la actuación reiterada de la policía, el ejército y otros funcionarios encargados de hacer cumplir la ley por violaciones graves de derechos humanos, como malos tratos y tortura (Amnistía Internacional, 2008).
- b) América. Amnistía Internacional (2008) señala a distintos países sudamericanos (por ejemplo Argentina, Ecuador, Brasil, Colombia o Cuba) por el trato que los funcionarios de la ley otorgan a las personas que se muestran disidentes con los organismos gubernamentales, especialmente en las cárceles y comisarías. Según esta organización, “la gran mayoría de las personas castigadas o encarceladas por los sistemas judiciales pertenecen a sectores desfavorecidos y no tienen recursos de ningún tipo. Los abusos de poder y los abusos contra los derechos humanos en los que se incluye la tortura suelen quedar impunes” (Amnistía Internacional, 2008, p. 25).

Además, el informe se presenta especialmente crítico con el comportamiento de los Estados Unidos a partir de los atentados del 11 de septiembre y la denominada “guerra contra el terrorismo internacional”:

Seis años después del comienzo de la denominada «guerra contra el terror», Estados Unidos seguía manteniendo recluidas bajo custodia militar indefinida a centenares de personas sin cargos ni juicio en Afganistán y Guantánamo, y a varios miles más en Irak. En julio, el presidente George W. Bush dio luz verde a la continuación del programa de detenciones e interrogatorios secretos puesto en marcha por la CIA. El hecho de que el presidente volviera a dar su visto bueno a este programa –una más de la larga lista de acciones políticas ilegales llevadas a cabo por el gobierno estadounidense en el marco de la «guerra contra el terror»– supuso un claro rechazo de los principios en los que se basa la Declaración Universal de Derechos Humanos. De hecho, el presidente Bush dictó su orden ejecutiva un año después de que dos órganos de la ONU encargados de la vigilancia de los tratados indicaran al gobierno estadounidense de manera inequívoca que las detenciones secretas constituían una violación de las obligaciones internacionales contraídas por Estados Unidos (Amnistía Internacional, 2008, pp. 25-26).

Respecto a las reticencias de los Estados Unidos a examinar las peticiones de *hábeas corpus* (derecho de los detenidos de que un juez dictamine si su detención fue legal o no) de los prisioneros de Guantánamo, Amnistía Internacional hace la siguiente denuncia:

El *hábeas corpus* –el derecho de una persona a que un juez determine la legalidad o no de su detención– es un principio fundamental de todo Estado de derecho. Las personas bajo custodia de Estados Unidos a las que se ha privado de este recurso se han convertido en objeto de desaparición forzada, traslado y detención secretos, procedimientos judiciales injustos, y tortura y otros tratos o penas crueles, inhumanos o degradantes. Sesenta años después de su adopción, este tipo de políticas y prácticas constituyen una afrenta al mundo que se deseaba construir con la Declaración Universal de Derechos Humanos (Amnistía Internacional, 2008, p. 26).

- c) Asia y Oceanía. “Asia continúa siendo la única región del mundo en donde no existe un instrumento general de derechos humanos” (Amnistía Internacional, 2008, p. 35). A partir de este planteamiento, el informe indica cómo las fuerzas de seguridad de distintos países cometen libremente violaciones de derechos humanos, como tortura y otros malos tratos. Veamos como ejemplo un fragmento de los datos que Amnistía Internacional presenta sobre China:

En junio, el Tribunal Supremo Popular estableció que los casos de pena de muerte correspondientes a tribunales de primera instancia debían juzgarse en sesiones públicas, y que los tribunales debían avanzar hacia la celebración de juicios públicos cuando se tratase de recursos presentados en casos de pena capital. Sin embargo, los juicios por delitos punibles con la pena de muerte continuaron celebrándose a puerta cerrada, la policía recurría a menudo a la tortura para obtener confesiones (Amnistía Internacional, 2008, p. 129).

- d) Europa y Asia Central. Si bien los casos de tortura y violaciones reiteradas de derechos humanos son mucho menores en los países europeos más desarrollados, sí se acusa a éstos de “colaborar” con programas de detenciones secretas e ilegales dirigidos por los Estados Unidos tras los atentados del 11 de Septiembre. Este tipo de detenciones implican generalmente tortura en los interrogatorios a los detenidos y otros malos tratos. Adicionalmente, el informe de Amnistía Internacional (2008) pone de relieve la pasividad de los gobiernos de algunos países (por ejemplo, Bosnia Herzegovina, España, Kazajistán, Moldavia, Reino Unido, Rusia, Turquía, Turkmenistán, Ucrania y Uzbekistán) a la hora de investigar casos de malos tratos y torturas, especialmente aquellos aplicados por la policía a inmigrantes detenidos o a refugiados políticos.
- e) Oriente medio y norte de África. Amnistía Internacional cita distintos países (por ejemplo, Argelia, Egipto, Túnez o distintos estados del Golfo Pérsico) en los que la tortura se emplea con relativa frecuencia como herramienta policial:

En toda la región, todopoderosos servicios de seguridad e inteligencia se ocupan de mantener el poder del Estado y reprimir la disidencia y el debate. Las personas que denuncian públicamente algo se exponen a ser detenidas arbitrariamente, recluidas sin juicio y sometidas a tortura y otros malos tratos por la policía de seguridad, cuyos señores políticos le permiten cometer abusos contra los derechos humanos con impunidad (Amnistía Internacional, 2008, p. 52).

En definitiva, pese a los distintos acuerdos y tratados internacionales que se han propuesto para acabar definitivamente con la tortura (véase Amnistía Internacional, 2008), ésta se sigue utilizando en la actualidad, en muchos casos para obtener información de sospechosos sobre crímenes o acciones contra el “partido” político que está en el poder.

A pesar de lo dicho, existe acuerdo entre los expertos contemporáneos acerca de la inutilidad de la tortura para obtener información fiable. Su problema fundamental reside, si dejamos de lado obvias consideraciones éticas y morales en torno a su empleo, en lo que Vrij (1998, en Masip, 2002) ha referido en relación con los puntos débiles del método de interrogatorio de Inbaud, Reid y Buckley (1986; Inbau, Reid, Buckley y Jane, 2001): que no hay ninguna técnica que haga que hablen solamente los culpables pero no los inocentes, sino que todo aquello que aumenta la probabilidad de que un culpable confiese, también aumenta la de que un inocente se autoincrimine falsamente. Es cierto que hoy la tortura se encuentra mucho menos extendida que en la mitad del siglo pasado, sobre todo en lo que se refiere a su admisión a nivel legal, pero a la luz de lo examinado anteriormente no sería sincero decir que realmente ha desaparecido.

Hasta este punto, se ha hecho un breve recorrido a lo largo de la historia con el fin de contemplar cómo las mentiras siempre han acompañado al hombre desde tiempos remotos hasta nuestros días. Así, las mentiras han estado presentes en los mitos de las principales creencias religiosas y corrientes filosóficas tradicionales. En estos casos, la mayoría de las veces se trataba de presentar a la mentira como una acción que reflejaba el “camino desviado” y “la traición” a la voluntad divina (desde la religión), o bien de reflejar el antagonismo entre como debe ser y no ser, comportarse y no comportarse, el hombre que busca la virtud y la sabiduría (desde la filosofía). A continuación se abordará una concepción más actual de la mentira, presentándose el tipo de mentiras consideradas en esta Tesis Doctoral. Es importante clarificar desde el principio con qué concepto y tipo de mentira se va a trabajar, dado que, como el lector puede intuir, no es sencillo establecer una única definición de “qué es mentir” que sea capaz de abarcar todas las connotaciones posibles y aplicarse en todos los contextos.

El objetivo en este apartado, pues, no es construir una nueva definición de mentira, ni discutir que estudioso de la mentira ha aportado una definición más correcta del término. De lo que se trata es de presentar el tipo específico de mentiras al que esta Tesis se refiere, así como las acotaciones que se han ido estableciendo para delimitarlo. En este recorrido, se observará un progresivo distanciamiento de la forma en que algunos estudiosos han abordado el concepto de mentira, así como de ciertos campos de estudio, no porque no se entiendan como adecuados dichos conceptos, tipos de mentira, o contextos de estudio, sino simplemente porque el tipo de mentira que se va a definir a continuación se refiere al considerado en este trabajo: las mentiras en el contexto jurídico legal.

EL PROBLEMA DE LA CONCEPTUALIZACIÓN

Definir qué es mentir no es una tarea fácil (Granhag y Strömwall, 2004). Tal como Masip, Garrido y Herrero (2004a) plantean, podría haber casi tantas definiciones de engaño como autores que lo han estudiado. El primer problema al que nos enfrentamos al tratar de definir el tipo de mentira al que se refiere esta Tesis, es un problema puramente lingüístico y derivado de dos potenciales aspectos que pueden dar lugar a confusión. En primer lugar, el hecho de que el español sea un idioma con abundantes sinónimos determina que no exista solamente un vocablo para referirse a la mentira. Por ejemplo, Gispert, Gay y Vidal (1996) han publicado un diccionario de sinónimos y antónimos en el que se incluyen más de 50 sinónimos de la palabra mentira. Algunos de ellos son: engaño, falsedad, embuste, añagaza, etc.

A lo largo de esta Tesis, de todos los vocablos posibles para referirse a la mentira se emplearán principalmente dos, mentira y engaño, y se utilizarán indistintamente⁴. La Real Academia Española define la mentira como la expresión o manifestación contraria de lo que se sabe, se cree o se piensa (Real Academia Española, 2004). Esta institución, a su vez, define el engaño como la falta de verdad en lo que se dice, hace, cree, piensa o discurre (Real Academia Española, 2004). Como veremos más adelante, las definiciones de los más importantes estudiosos de la mentira, desde una perspectiva más psicológica que lingüística o filológica, matizarán en mayor o menor grado estas definiciones dependiendo de la perspectiva teórica en la que se sitúe el autor. Pero la presentación del concepto tal como lo define la Real Academia Española permite observar la equivalencia de los dos términos.

⁴ La razón por la que hemos optado por utilizar indistintamente los términos mentira y engaño y no otros, es porque son los mayormente empleados en las distintas publicaciones científicas sobre el tema, tanto en los trabajos publicados en español, como en el caso de sus vocablos equivalentes en las publicaciones en inglés (*lie* y *deception*). No obstante, algunos estudiosos como Hopper y Bell (1984) o posteriormente Bok (1978, en Meltzer, 2003) han empleado específicamente el término mentira (*lie*), para referirse específicamente a mentiras verbales y no de otro tipo. Actualmente, en la literatura científica la tendencia predominante es emplear el término indistintamente, tanto para mentiras verbales como para las no verbales. Así pues, al igual que en trabajos dedicados específicamente a la conceptualización del término (véase, por ejemplo, Masip, Garrido y Herrero, 2004a), emplearemos de forma equivalente los vocablos mentira y engaño.

Un segundo aspecto que podría ser problemático es que los estudios de la mentira no son patrimonio exclusivo de ningún idioma. Actualmente, la mayoría de las publicaciones más relevantes sobre la mentira se imprimen en lengua inglesa, y en algunos casos por autores de habla no inglesa pero que publican sus trabajos en inglés. De esta forma, y a pesar de que el inglés es un idioma cada vez más global, es común la circunstancia de que un investigador de habla no inglesa ni española (por ejemplo alemán) publica en inglés, y esta publicación la lee un lector español. De esta forma, una publicación podría ser conceptualizada en alemán, escrita en inglés, y leída por un español. Siguiendo nuestro ejemplo, un lector español tendría que traducir al español los términos de la lengua inglesa para referirse a la mentira que un autor alemán previamente tradujo del alemán al inglés para escribir su publicación. Si bien es cierto que actualmente no sólo se publica en inglés, sino que ya directamente en muchos casos se analizan y discuten los datos y resultados en ésta lengua, también es cierto que los datos originalmente obtenidos (por ejemplo, respuestas a cuestionarios) suelen estar en la lengua original del país en el que se recogen, y en el proceso de traducción posterior al inglés podrían perderse o tergiversarse connotaciones sutiles de significado. Consultando el *Collins Spanish Dictionary* (Evans y Harwood, 1991) se observa que la palabra inglesa para nombrar la mentira es *lie*, mientras que para referirse al engaño es *deception*.

EL TIPO DE MENTIRA CONSIDERADO EN ESTA TESIS

Distintos autores han tratado de delimitar las características que debe tener una comunicación para ser considerada una mentira, lo cual ha dado lugar a definiciones que hacen hincapié en aspectos diferentes. Por ejemplo, Coleman y Kay (1981) señalan tres elementos definitorios de una mentira: (a) la creencia del emisor en la falsedad de la afirmación, (b) la intención de mentir, y (c) la falsedad objetiva de la información transmitida.

Teniendo en cuenta que esta Tesis se enmarca en un contexto muy específico, la mentira en contextos jurídico legales, se presentan a continuación los cuatro requisitos que se han considerado para acotar de la forma más

precisa posible el tipo de mentiras con las que se ha trabajado. Como veremos, algunos de ellos coinciden con los propuestos por Coleman y Kay (1981).

En concreto, esta Tesis se ha centrado exclusivamente en aquellas mentiras que se producen en un contexto de interacción entre al menos dos personas, en el que una persona A (quien ejecuta la conducta mendaz) trata (con o sin éxito) de generar en la mente de una persona B una idea que cree que no es cierta, a través de su conducta verbal o no verbal. La persona mendaz A, además, actúa consciente y deliberadamente, y sin haber advertido previamente a la persona B acerca de su intención. A continuación se analizan cuatro aspectos clave de este tipo de mentiras.

1. MENTIRAS ENTRE SERES HUMANOS

La primera acotación establecida al plantear esta definición, se refiere al tipo de “actores” que van a participar en los procesos de engaño: personas. Algunos investigadores han planteado la posibilidad de estudiar y comprender la mentira como una conducta que se ha desarrollado evolutivamente, y que posee un importante valor adaptativo (Anderson, 1986; Bond, Kahler y Paolicelli, 1985; Chevalier, 1986; Menzel, 1986; Mitchell, 1986). Los trabajos de estos autores ofrecen interesantes perspectivas desde las cuales podría plantearse la posibilidad de que el engaño no sea un acto exclusivamente humano, sino que también pueda darse a distintos niveles en otras especies animales. Resultaría muy interesante entrar en la discusión de si, efectivamente, se puede hablar de mentiras entre animales, o entre animales y humanos, y en tal caso, qué características tendría que tener la definición a considerar y a qué niveles se podría hablar de engaño. No obstante, esa discusión se apartaría del objetivo de esta Tesis y del tipo específico de mentiras que hemos definido. Así pues, en este trabajo nos referiremos exclusivamente a las mentiras que se dan entre seres humanos.

2. MENTIRAS DIRIGIDAS A OTRAS PERSONAS

El siguiente aspecto en la acotación del tipo de mentira que se ha definido se refiere al hecho de que se considerarán exclusivamente aquellas mentiras con

las que una persona trata de engañar a otra persona distinta. Es decir, mentiras que ocurren entre, al menos, dos personas.

Esta delimitación es interesante en la medida en que algunos estudiosos del engaño han planteado la existencia del “autoengaño” como un proceso mediante el cual un individuo se miente a sí mismo. Por ejemplo, Ekman (1997a, 1997b) propone que el autoengaño ocurre cuando una persona se está engañando a sí misma sin ser consciente de ello y sin saber los propios motivos del engaño. El autoengaño, pues, no entraría en la definición planteada al principio de este apartado, y por lo tanto dentro del tipo de mentiras estudiadas en esta Tesis, por dos razones principalmente: (a) en primer lugar, porque en el autoengaño solamente participa una persona, y (b) porque, según la definición adoptada, hay un requisito que no está claro que cumpla el autoengaño, y que se refiere al hecho de que la persona mendaz debe actuar conscientemente.

3. INTENCIONALIDAD

El hecho de que en el momento de mentir el individuo debe ejecutar tal conducta intencionalmente es un aspecto clave en la mayoría de las definiciones de la mentira existentes. Por ejemplo, Krauss (1981, en Vrij, 2000) define el engaño como un acto por el que la persona que miente pretende crear en otra persona una creencia o comprensión de algo que el propio comunicador (mentiroso) considera que es falso. Esta “pretensión” u objetivo (crear una idea falsa), refleja la idea de la intencionalidad del comunicador (mentiroso), que deliberadamente realizará las conductas que considere más apropiadas para alcanzar su meta. Posteriormente, otros autores irán completando esta definición añadiendo otros matices, pero conservando siempre la idea de que para poder hablar de mentiras, es necesario que la persona que miente tenga realmente el propósito de hacerlo.

En la definición de engaño propuesta por Vrij (2000) aparece también la idea de la intencionalidad del individuo que miente, al calificar la conducta como deliberada:

Un intento deliberado (exitoso o no) de generar en otra persona una creencia que el propio comunicador considera falsa sin advertir previamente al receptor de la comunicación (Vrij, 2000, p. 6).

Posteriormente Masip (2002) y después Masip et al. (2004a) presentan una de las definiciones más completas del concepto de engaño de cuantas se han ofrecido:

Un intento deliberado, exitoso o no, de ocultar, fabricar y/o manipular de cualquier otra forma información factual y/o emocional mediante su conducta verbal y/o no verbal, con el objetivo de crear o mantener en otro u otros una creencia que el propio comunicador considera falsa (Masip et al., 2004a, p. 147).

En esta definición, se presenta en primer lugar la idea de la intencionalidad del mentiroso en la medida en que se trata de un intento deliberado. Es el propio mentiroso quien deliberadamente decide mentir, quien tiene la intención de mentir. En conclusión, tal como plantean Granhag y Strömwall (2004) y Masip et al. (2004a) la mentira requiere intencionalidad. Estos autores plantean además que una persona que sin intención ofrece información que no se corresponde con la realidad (por ejemplo, recuerdos erróneos de hechos acontecidos en el pasado), no está mintiendo, precisamente porque no es esa su intención. Se podría quizá hablar de información errónea, pero no de mentira. Así pues, intencionalidad, y por consiguiente consciencia, son requisitos imprescindibles para poder hablar de mentira.

4. TRANSMISIÓN DE UNA CREENCIA QUE EL MENTIROSO CONSIDERA FALSA

Finalmente, tal como refleja la definición de Masip et al. (2004a), la creencia del mentiroso de que la información que va a transmitir es falsa es también un requisito imprescindible para que una comunicación sea considerada mentira.

Para poder afirmar que una información transmitida es una mentira, es necesario que el propio comunicador la crea falsa. Si el emisor considera verdadera una información pero ésta no se corresponde con la realidad objetiva, podría hablarse de un error o confusión por parte del comunicador, pero no de mentira, ya que falta el componente intencional. Por otra parte, un

emisor que estuviera dando información correcta que él considera falsa estaría mintiendo. En este sentido, aunque Coleman y Kay (1981) indican la creencia del emisor en la falsedad de la afirmación y la falsedad objetiva de la misma, como dos criterios definitorios de la mentira, el primero de ellos cobra sin duda mayor importancia que el segundo.

En definitiva, como se ha indicado al principio de este apartado, el objetivo no era establecer un nuevo concepto de “qué es mentir”, ni discutir las distintas definiciones que distintos autores han propuesto, sino simplemente clarificar el tipo de mentiras que se han considerado en esta Tesis. Así pues, a lo largo del presente trabajo nos referiremos exclusivamente a mentiras (a) emitidas por seres humanos, (b) dirigidas a otros seres humanos, (c) que se cuentan consciente e intencionalmente, y (d) en las que el comunicador cuenta algo que cree que es falso.

PREVALENCIA DE LA MENTIRA EN LAS RELACIONES HUMANAS.

Es tan cierto que los más importantes sistemas filosóficos, teológicos, legislativos, éticos y morales, históricamente han ofrecido un rechazo y una condena radical de la mentira, como que actualmente, y por mucho que a algunos les cueste admitirlo, la mentira es una realidad inevitable de nuestra vida.

Tal como plantean diversos autores, se puede considerar el engaño como un tipo de comunicación que, en determinados contextos y situaciones, es deseable no solamente para el mentiroso, sino también para el destinatario de la mentira, para terceras personas, o para preservar la convivencia y las relaciones personales (Buller y Burgoon, 1994; DePaulo, 2004; Knapp, Hart y Dennis, 1974). Como veremos a continuación, y a pesar de imperativos religioso-morales tan radicales como utópicos, mentimos más de lo que pensamos que lo hacemos. Mentimos no solamente para obtener recompensas sociales, sino también para afianzar nuestras relaciones, para sentirnos seguros de nosotros mismos, para proteger nuestra autoestima, para proteger el autoconcepto y el concepto que tenemos de los demás, etc.

Distintos estudios han tratado de establecer la prevalencia del engaño en las relaciones sociales. Por ejemplo, Turner, Edgley y Olmstead (1975, en O'Hair y Cody, 1994) encontraron que aproximadamente el 62% de las aserciones conversacionales emitidas por los participantes de su estudio podían ser clasificadas como mentiras. Venant (1991, en O'Hair y Cody, 1994) obtuvo que un 97% de su muestra de 5700 participantes admitió haber mentido en sus relaciones sociales, y una tercera parte de aquellos que estaban casados indicaron también haber mentido a sus esposas. DePaulo, Kashy, Kirkendol, Wyer y Epstein (1996) realizaron uno de los estudios más reveladores respecto a la prevalencia de las mentiras en la vida diaria. Los participantes fueron 77 estudiantes de edades comprendidas entre 17 y 22 años ($M = 18.69$ años, $DE = 0.91$), y 70 miembros de una comunidad residencial, en su mayoría trabajadores de edades comprendidas entre 18 y 71 años ($M = 34.19$ años, $DE = 12.49$). La tarea de los participantes consistía en registrar durante una semana todas las interacciones que tuviesen con cualquier otra persona y todas las mentiras que contaran durante dichas interacciones. DePaulo et al. (1996) encontraron que los estudiantes mentían una media de dos veces al día y los miembros de la comunidad residencial una media de una vez al día.

Teniendo en cuenta, pues, que el acto de mentir tiene un carácter más cotidiano que extraordinario, las preguntas que se suscitan a continuación son: ¿sobre qué se miente más a menudo?, ¿cuáles son las razones más comunes para mentir?, ¿qué tipo de mentiras son las más habituales? En el estudio anteriormente mencionado, DePaulo et al. (1996) responden algunas de estas preguntas. En la Tabla 1 se presentan algunos de los resultados de estos autores. A partir de los datos de esta tabla concluyen que: (a) tanto los estudiantes como los miembros de la comunidad residencial mentían sobre lo mismo; (b) ambos grupos mentían más frecuentemente sobre sentimientos y sobre acciones y planes que sobre otros aspectos; (c) los dos grupos de participantes mentían más por razones centradas en uno mismo que por razones centradas en otros; y (d) el tipo de mentiras más frecuentes entre los integrantes de los dos grupos fueron las falsedades totales.

Tabla 1. *Contenido, Razones y Tipos de Mentiras Expresados en Porcentajes Medios (DePaulo et al., 1996).*

	Estudiantes		Miembros de comunidad residencial	
	M	DE	M	DE
<i>Contenido de las mentiras</i>				
Sentimientos	37.42	25.93	29.53	29.97
Conseguir logros	15.84	16.87	17.14	24.89
Acciones, planes	27.49	25.57	27.69	27.27
Explicaciones	10.28	14.76	11.17	15.49
Hechos, posesiones	8.97	12.99	14.55	23.46
<i>Razones para mentir</i>				
Centradas en uno mismo ^a	45.48	27.63	56.68	31.24
Centradas en otros ^b	25.74	24.42	24.45	27.20
<i>Tipos de mentiras</i>				
Falsedades totales ^c	67.63	24.85	59.18	33.28
Exageraciones ^d	14.74	19.30	9.23	17.80
Mentiras sutiles ^e	8.62	13.18	23.19	28.48

^aMentiras dichas para protegerse uno mismo u ofrecer algún tipo de beneficio.

^bMentiras dichas para proteger u ofrecer algún tipo de beneficio a otros

^cEn inglés *outright lies*. Mentiras en las que la información mostrada es totalmente diferente o contradictoria con la verdad.

^dMentiras que se producen por sobreestimación de los hechos, excediendo así la verdad.

^eEl término es *subtle*. Se refiere a aquellos casos en los que la información que se cuenta es cierta, pero utilizando un lenguaje, tanto verbal o no verbal, con el que se pretende que dicha información no sea creída.

Si bien DePaulo et al. (1996) estudiaron la prevalencia de las mentiras en todo tipo de interacciones sociales durante un periodo de una semana, otros estudios se han centrado específicamente en las mentiras en las relaciones íntimas de pareja, observando resultados similares. Por ejemplo, Knox, Schacht, Holt y Turner (1993, en Cole, 2001) encontraron que el 92% de los participantes de su estudio admitió que había mentido a su pareja.

Hasta el momento, se ha observado que distintos estudios parecen apuntar en la dirección de que el acto de mentir es una conducta relativamente frecuente en las interacciones cotidianas. Centrándonos ahora en la persona que miente, podríamos plantear las siguientes preguntas: ¿Cómo es la gente que más miente?, ¿existen diferencias entre aquellos que mienten más a menudo y quienes lo hacen menos frecuentemente? Kashy y DePaulo (1996) realizaron un interesante estudio en el que trataron de establecer una correlación entre distintos aspectos de personalidad y la cantidad y tipo de mentiras contadas, bien para conseguir beneficios para uno mismo o para los demás. Las autoras hallaron que los individuos que más mienten son aquellos más preocupados por crear una buena impresión en sus relaciones sociales, más extrovertidos que introvertidos, y más manipuladores, aunque, tal como

indican las propias autoras, “delicada e incluso deseablemente manipuladores, más que abrasivamente manipuladores” (Kashy y DePaulo, 1996, p. 1048). En cierto sentido, según los resultados de este estudio, serían individuos bien capacitados para una vida social activa. No obstante, si bien es cierto que tienen más relaciones sociales, también parece que viven estas relaciones de forma menos gratificante. Además, suele tratarse de personas que, en cierto modo, tienen un menor nivel de socialización; es decir, tienen menos interiorizados los valores tradicionales propios de su cultura (en este caso la cultura occidental).

Además, parece que las personas menos manipuladoras, más altamente socializadas y con mejor calidad de relaciones sociales suelen contar más mentiras orientadas al beneficio de otros que a su propio beneficio.

No queremos terminar este apartado sin hacer una referencia a las mentiras *on-line*, y es que las nuevas tecnologías tampoco se libran de las mentiras. George et al. (2004), en la *International Conference on System Sciences 2004*, han hecho referencia al aumento cada vez mayor de las mentiras y engaños en Internet, especialmente a través de aquellos *e-mails* en los que, de alguna forma, se trata de obtener algún beneficio de los receptores más confiados. Hasta tal punto están proliferando estos mensajes, que se han creado páginas web especializadas solamente en proporcionar sistemas de detección e información sobre los mismos.

En conclusión, teniendo en cuenta todo lo anterior y aunque nos cueste admitirlo, en la literatura del engaño hay un acuerdo general sobre que la mentira cotidiana es un fenómeno generalizado en nuestras relaciones sociales. Un fenómeno que, tal como algunos autores defienden, es incluso necesario y adaptativo (DePaulo y Kashy, 1998; DePaulo et al., 1996; DePaulo, Wetzel, Weylin y Walter, 2003; Doner, 1998; Ford, 2006; Geary, 2000; Meltzer, 2003; O’Hair y Cody, 1994; Seiter, Brusckke y Chunsheng, 2002; Tornqvist, Anderson y DePaulo, 2001; Vrij, 2000; Zuckerman, DePaulo y Rosenthal, 1981).

TRES VIAS PARA DETECTAR MENTIRAS

Actualmente, el estudio científico de la detección del engaño distingue tres aproximaciones distintas desde las que se puede enfocar la difícil tarea de tratar de saber cuándo alguien está mintiendo a partir de su conducta (véase, por ejemplo, Granhag y Vrij, 2005; Masip y Garrido, 1999, 2000, 2001a; Masip et al., 2002a), en concreto: (a) la aproximación basada en indicadores conductuales no verbales, (b) la aproximación basada en el análisis del contenido verbal de la declaración, y (c) la aproximación basada en indicadores psicofisiológicos. En este apartado describiremos brevemente cada uno de estos enfoques.

APROXIMACIÓN BASADA EN INDICADORES CONDUCTUALES NO VERBALES

¿Existen razones para pensar en la existencia de indicios no verbales de engaño?

Una de las cuestiones principales que ha guiado la investigación desde esta aproximación es si se puede detectar una mentira a través de la conducta no verbal (visible o paralingüística) del individuo. Se presentan a continuación los planteamientos de una de las teorías más clásicas al respecto: la teoría de los cuatro factores. Propuesta por Zuckerman et al. (1981), esta teoría sostiene que habría cuatro causas principales por las que una persona, cuando miente, puede mostrar indicios conductuales observables que pueden delatarle:

a) Intento de control. Cuando un individuo miente, puede intentar controlar su conducta tratando de evitar determinados movimientos o gestos que considera que pueden indicar que está mintiendo. Sin embargo, según Zuckerman et al. (1981), este sobre-control puede dar lugar a que la conducta parezca planeada, previamente ensayada y carente de espontaneidad.

Además, pueden aparecer discrepancias entre aquellas conductas fácilmente controlables y otras más automáticas y difíciles de ejecutar o inhibir a voluntad.

b) *Arousal*. Basándose en la idea de que al mentir experimentamos más *arousal* que cuando decimos la verdad, se plantea que este aumento puede traducirse en la aparición de ciertos indicios cuya presencia puede ser signo de que no se está siendo sincero. Algunos de tales indicios pueden ser: el aumento del tono de voz, la dilatación pupilar o los pestañeos.

c) *Afecto*. Como señala Ekman (1985/1992), las emociones más comúnmente asociadas con el acto de mentir son la culpa por implicarse en este tipo de conductas, la ansiedad por la posibilidad de ser descubierto, y lo que se ha llamado placer de engañar (sensación de deleite o exaltación por la acción de engañar a alguien). Zuckerman et al. (1981) plantean que al mentir van a aparecer una serie de indicios derivados de estas emociones que pueden delatar al mentiroso (véase también Ekman, 1985/1992).

d) *Sobrecarga cognitiva*. Mentir es una tarea más compleja que decir la verdad. Dicha complejidad va a exigir un sobre-esfuerzo cognitivo que se va a reflejar en la aparición de conductas no verbales específicas observables (véase, por ejemplo, Vrij, 2000; Vrij, Fisher, Mann y Leal, 2006; Vrij et al., 2008). Algunas de estas conductas serían un decremento de movimientos de los pies y las piernas, más errores al hablar o una mayor lentitud al hablar.

La Teoría de los cuatro factores representa uno de los fundamentos teóricos más importantes para pensar que la conducta al mentir difiere de la conducta al decir la verdad, y sus planteamientos han supuesto la base teórica a partir de la que se han realizado muchos de los estudios que describiremos a continuación.

La investigación

Hacer un repaso exhaustivo de cómo se ha ido desarrollando la investigación en detección del engaño a nivel no verbal desde sus inicios hasta nuestros días sobrepasa el objetivo de esta Tesis. Tampoco se pretende en este trabajo desarrollar una propuesta novedosa de clasificación de los

estudios publicados. Simplemente, y con ánimo de no alargar en exceso este trabajo, nos limitaremos a exponer los resultados de las principales investigaciones respecto a los tres interrogantes probablemente más importantes que se podrían plantear desde esta perspectiva: (a) ¿Existen realmente indicios no verbales eficaces para detectar el engaño?, (b) ¿cuáles son los indicios en los que los individuos habitualmente basan sus juicios?, y (c) ¿cual es la precisión (nivel de aciertos) en la detección de mentiras a partir de la observación de la conducta no verbal?

A. ¿Existen indicios no verbales realmente eficaces para detectar el engaño?

Ninguno de los estudios realizados, desde los inicios de la investigación hasta la más reciente publicación, ha encontrado un indicio o conducta no verbal que indique con seguridad absoluta que un individuo está mintiendo o diciendo la verdad. Sin embargo, sí es cierto que cuando una persona miente es más probable que experimente ciertos procesos psicológicos que probablemente den lugar a la aparición de conductas observables. La relación, pues, que se establece entre la aparición de algunas conductas no verbales y la posibilidad de que el individuo que las muestra esté mintiendo no es de carácter absoluto, sino probabilístico (por ejemplo, DePaulo, Lindsay, et al., 2003; DePaulo y Morris, 2004; DePaulo, Zuckerman y Rosenthal 1980; Granhag y Vrij, 2005; Vrij, 2000).

Teniendo en cuenta lo anterior, distintas investigaciones han tratado de delimitar qué conductas no verbales y en qué medida se relacionan realmente con el engaño. En la Tabla 2 aparecen resumidos los resultados de los meta-análisis y revisiones más importantes. Observando esta tabla, lo más llamativo es que no hay un claro acuerdo entre las distintas revisiones sobre qué indicios no verbales están realmente asociados al engaño y en qué sentido. Con excepción de la dilatación pupilar y el tono de voz, cuyo aumento se ha encontrado significativamente asociado al acto de mentir en todos los meta-análisis en los que se han analizado, no se observa un patrón claro para los demás indicios. No obstante, aunque la Tabla 2 ofrece una visión global, no sería adecuado sopesar los resultados de todas las revisiones y meta-análisis presentados por igual, ya que ni la cantidad de estudios que cada uno ha

contemplado, ni la forma de analizar los datos ha sido igual en todos los casos. En este sentido, se expondrá a continuación de forma más detallada el trabajo de DePaulo, Lindsay, et al. (2003) por ser el más inclusivo en cuanto a estudios analizados e indicios examinados, y los meta-análisis de Sporer y Schwandt (2006, 2007) por ser los más recientes.

Meta-análisis de DePaulo, Lindsay, et al. (2003)

DePaulo, Lindsay, et al. (2003), desde una perspectiva auto-presentacional⁵ de la mentira (expuesta ya previamente en trabajos como DePaulo, 1992), realizaron el más extenso meta-análisis publicado hasta el momento sobre indicios de engaño. En el estudio se incluyó un total de 120 muestras de participantes extraídas de distintas investigaciones, 158 indicios (97 de ellos no verbales) y 1338 estimaciones del vínculo entre la ocurrencia de un indicio y la mentira.

Los autores utilizaron el estadístico d para calcular el tamaño del efecto de cada indicio. Dicho estadístico se obtuvo restando la media de la aparición del indicio correspondiente al decir la verdad de la media de la aparición de dicho indicio al mentir, dividiendo entonces el resultado entre la media de las desviaciones típicas correspondientes. De esta forma, valores positivos de d indicarían que el indicio analizado aparece más al mentir, mientras que valores negativos significarían que dicha conducta aparece más al decir la verdad. Finalmente, valores de d cercanos a cero indicarían una falta de relación entre el indicio y mentir o decir la verdad.

⁵ Desde la perspectiva auto-presentacional o teoría del manejo de la impresión, se plantea esencialmente que los individuos regulan su conducta con fines auto-presentacionales; es decir, para ofrecer a otras personas una u otra imagen según les convenga. En este sentido, cuando un individuo desea ser creído ajustará su conducta en función de cómo él piensa que se comporta alguien que dice la verdad. De esta forma es probable que una persona que miente, en su empeño por parecer sincera muestre un comportamiento exagerado, tanto respecto a la inclusión de ciertas conductas o elementos (que considera que reforzarán su credibilidad) como a la supresión de otros (que considera signos de mentira) (véase, por ejemplo, DePaulo, 1992, para una explicación más detallada de los postulados de esta teoría).

Tabla 2. Relación de Distintos Indicios no Verbales con el Engaño Según los Resultados de Distintas Revisiones y Meta-análisis.

Indicios	Estudios						
	Kraut (1980)	Zuckerman et al. (1981)	Vrij (2000)	DePaulo, Lindsay, et al. (2003)	Sporer y Schwandt (2006)		Sporer y Schwandt (2007) ^a
					Datos ponderados	Datos sin ponderar	
Adaptadores (en general)	>	>	-	>			-
Apretar los labios				>			
Asentir con la cabeza				-			<
Bienestar facial ^b				<			
Cooperatividad				<			
Dilatación pupilar		>		>			
Discrepancias, ambivalencias				>			
Duración del discurso		-		<	-	<	
Duración de las pausas			>				
Elevación del mentón				>			
Encoger los hombros		>		-			
Errores del habla	>	>	-	-	-	>	
Ilustradores				<			-
Impresiones de incertidumbre verbal y vocal				>			
Impresiones de inmediatez verbal y vocal				<			
Implicación verbal y vocal				<			
Latencia de respuesta		-	-	-	>	>	
Movimientos de brazos y manos			<	-			
Movimientos sólo de manos				-			<
Movimientos de piernas y pies		-	<	-			<
Nerviosismo (general)				>			
Parpadeos	>	-	-	-			-
Tensión vocal				>			
Titubeos / vacilaciones	>	>	-	-			
Tono de voz		>	>	>	>	>	

Nota. Se utiliza (>) para indicar que la frecuencia de aparición del indicio aumenta al mentir; (<) para indicar que disminuye, y (-) para indicar que no se encontró relación.

^aLos autores presentan solamente los datos ponderados dada la equivalencia con los datos sin ponderar.

^bEl término original en inglés es *facial pleasantness*.

DePaulo, Lindsay, et al. (2003) organizaron los resultados referidos al valor diagnóstico de cada indicio en función del número de estudios en que se había examinado dicho indicio (más o menos de cinco estudios) y de su tamaño del efecto (se consideró que valores de d superiores a .20 o inferiores a -.20 eran indicativos de que el indicio sería una herramienta adecuada para detectar mentiras; aunque otras conductas discriminaron significativamente ($p < .05$), los autores no las consideraron adecuadas si el tamaño del efecto no era superior a .20 o inferior a -.20). En las Tablas 3, 4 y 5 aparece un resumen de los resultados más significativos a nivel global. DePaulo, Lindsay, et al. (2003) estudiaron 97 indicios no verbales, pero solamente 37 se habían examinado en más de cinco estudios. Adicionalmente, sólo siete de estos últimos tuvieron un tamaño del efecto mayor que .20 o menor que -.20 (ver Tabla 3). De esta forma, los indicios que cumplieron los requisitos propuestos para ser considerados una buena herramienta de discriminación fueron muy pocos (Tabla 3).

Otros 22 indicios no verbales se examinaron tan solo en tres, cuatro o cinco estudios. De estos, únicamente cinco tuvieron un tamaño del efecto mayor que .20 o menor que -.20 (Tabla 4), con lo que, aunque aparentemente podrían considerarse una buena herramienta de discriminación, se necesita más evidencia experimental al respecto.

Finalmente, hubo un total de ocho indicios que, aunque tuvieron un tamaño del efecto mayor que .20 o menor que -.20, solamente se pudieron examinar en uno o dos estudios. Dada la acusada falta de evidencia experimental, nos limitaremos solamente a presentarlos, pero no se considerarán en una discusión posterior (ver Tabla 5).

Tabla 3. *Indicios con $d > .20$ ó $d < -.20$ Examinados en más de Cinco Estudios que Mostraron una Relación Significativa con la Mentira.*

Indicio	Definición	<i>d</i>	Estudios en que se examinó
Discrepancia, ambivalencia	La comunicación del hablante parece internamente discrepante o inconsistente. La información de distintas fuentes (ej. cara vs. cuerpo) parece contradictoria; el hablante parece ambivalente.	.34	7
Frecuencia, tono	El tono o frecuencia de la voz suena alto.	.21	12
Implicación verbal y vocal	El hablante describe experiencias personales o acontecimientos de forma personal y reveladora. El hablante parece expresivo e implicado a nivel vocal.	-.21	7
Impresiones de inmediatez vocal y verbal	El hablante contesta de forma que parece directo, relevante, claro y personal en vez de indirecto, distante, evasivo, irrelevante, poco claro o impersonal.	-.55	7
Impresiones de incertidumbre verbal y vocal	El hablante muestra incertidumbre, inseguridad, poco dominio, asertividad o énfasis. El hablante parece tener dificultad en contestar a la pregunta.	.30	10
Nerviosismo, tensión (general)	El hablante parece nervioso, tenso, hace movimientos corporales que parecen nerviosos.	.27	16
Tensión vocal	La voz suena tensa, no relajada, o con estrés vocal.	.26	10

Tabla 4. *Indicios con $d > .20$ ó $d < -.20$ Examinados en Cinco, Cuatro o Tres Estudios que Mostraron una Relación Significativa con la Mentira.*

Indicio	Definición	<i>d</i>	Estudios en que se examinó
Cooperativo (general)	El hablante parece cooperativo, dispuesto a ayudar, positivo y seguro.	-.66	3
Dilatación pupilar	Aumento del tamaño de la pupila (usualmente medido con un pupilómetro).	.39	4
Elevación del mentón	El mentón aparece elevado. El labio inferior y el mentón están flexionados.	.25	4
Repeticiones de palabras y frases	Las palabras o frases se repiten sin que aparezcan pausas o errores del lenguaje.	.21	4
Tiempo hablando	Proporción del tiempo total que el declarante pasa hablando o parece hablador.	-.35	4

Tabla 5. *Indicios con $d > .20$ ó $d < -.20$ Examinados en uno o dos Estudios que Mostraron una Relación Significativa con la Mentira.*

Indicio	Definición	<i>d</i>	Estudios en que se examinó
Cambios en los movimientos de los pies	Cambios en el número movimientos de pies y piernas a lo largo del tiempo.	1.05	2
Cambios pupilares	Cambios en el tamaño de la pupila.	.90	1
Falta de espontaneidad del discurso, parece planeado	El mensaje parece planeado o ensayado.	.35	1
Indiferencia, despreocupación	El hablante parece indiferente, despreocupado.	.59	2
Intensidad de la expresión facial	La expresión facial del hablante parece intensa.	-.32	2
Movimientos específicos de manos y dedos	Movimientos de las manos que no incluyen movimientos de los brazos, y movimientos de los dedos que no incluyen movimientos de las manos.	-.36	2
Palabras interrumpidas y repetidas	Palabras interrumpidas y repetidas (subconjunto específico de la categoría interrupciones).	.38	1
Sonrisa genuina	Definida por el movimiento de los músculos <i>orbicularis oculi</i> (que rodea el ojo) y zigomático mayor durante emociones positivas.	-.70	2

Observando las 12 conductas no verbales que mostraron adecuados tamaños de efecto en al menos tres estudios (Tablas 3 y 4), se aprecia que tres de ellas (dilatación pupilar, tono de voz y tensión vocal) son de muy difícil observación directa, es decir, sin la ayuda de instrumentos adecuados. De los indicios restantes, seis (nerviosismo (general), cooperatividad, impresiones de incertidumbre verbal y vocal, impresiones de inmediatez verbal y vocal, implicación verbal y vocal, y discrepancia-ambivalencia) son un tanto complejos de identificar y vagos a la hora de definir. Por ejemplo, respecto a la impresión general de nerviosismo, tal como los autores definen esta conducta no está claro cuándo se puede concluir su ocurrencia. Por un lado se hace referencia a movimientos corporales que parecen nerviosos. Habría, pues, que atender a distintos movimientos corporales. ¿Cuáles y cuantos exactamente? En la definición no se especifica. Por otro lado, no está claro cómo deben ser dichos

movimientos. ¿Cómo tiene que ser la intensidad y frecuencia de cada uno de ellos? Los autores utilizan la expresión “general”, lo que deja la puerta abierta a distintas interpretaciones de distintos observadores de una misma conducta. No hay una delimitación conceptual precisa del indicio. Lo mismo podríamos decir de la cooperatividad “general”. Con las impresiones de inmediatez verbal y vocal y las impresiones de incertidumbre verbal y vocal nos encontramos ante un problema similar. En ambos casos los autores se refieren a “impresiones”, que ya de por sí es un término vago. Además, por ejemplo, respecto a la inmediatez (ver Tabla 3), ¿cómo debe ser exactamente una respuesta para parecer directa, relevante y personal? No se especifica. ¿Dónde está el límite, por ejemplo, entre una respuesta directa o indirecta, personal o impersonal? La definición introduce el término “parece”, que es subjetivo, y no establece un criterio claro y general de diferenciación. Con las impresiones de incertidumbre ocurre lo mismo. Tampoco está claramente definida la implicación verbal y vocal. De nuevo se utilizan términos ambiguos (ver Tabla 3) como “parece expresivo e implicado” que dejan la interpretación de la conducta del emisor a la subjetividad de cada observador. Respecto a la discrepancia-ambivalencia, se trata de un indicio que implica un doble proceso para ser identificado. Por un lado el observador debe ser capaz de codificar simultáneamente información de dos o más fuentes; por ejemplo, el rostro y el cuerpo. Además, el observador debe establecer una comparación entre los indicios ofrecidos por dichas fuentes y ver si son discrepantes entre sí, todo ello siguiendo el curso normal de una conversación. Además, sería conveniente comparar los indicios de los distintos canales (el rostro y el cuerpo) con el contenido del discurso para saber a través de qué canal exactamente se emite la conducta discrepante. Por ejemplo, el declarante puede estar describiendo una experiencia calificada por él como muy negativa, mostrar una conducta corporal acorde con la declaración, pero su rostro puede filtrar indicios de júbilo. En este caso habría una discrepancia entre el rostro y el cuerpo, siendo el rostro la fuente de indicios discrepantes con la declaración. En definitiva, aquellos indicios no verbales que se han mostrado realmente asociados a la mentira son, en ocasiones, conductas cuya observación no es relativamente sencilla.

Finalmente, DePaulo, Lindsay, et al. (2003) descubrieron que existen distintas variables moderadoras que hacen que ciertos indicios discriminen bajo ciertas condiciones pero no bajo otras. A continuación se describen los resultados más significativos sobre el efecto de dichos moderadores.

a) La motivación para mentir con éxito. De los 13 indicios en los que se pudo comparar si había diferencias en significatividad en función de una alta o baja motivación para mentir, en cuatro de ellos se observaron cambios significativos producidos por esta variable moderadora.

Tabla 6. *Tamaño del Efecto (d) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función de la Motivación (DePaulo, Lindsay, et al., 2003).*

	Sin motivación	Con motivación
Contacto ocular	.09	-.15*
Movimientos de pies o piernas	-.02	-.13*
Nerviosismo, tensión	.15	.35*
Frecuencia/tono de voz	-.02	.59*

* $p < .05$.

Los cuatro indicios que aparecen en la Tabla 6 no discriminaron significativamente verdades de mentiras en aquellos estudios en los que los emisores no tenían una motivación especial para mentir con éxito. Sin embargo, los resultados de las investigaciones en las que los participantes sí estaban motivados, muestran una discriminación significativa para todos ellos. En condiciones de alta motivación, el contacto ocular y los movimientos de pies o piernas disminuyeron significativamente al mentir, mientras que la tensión y el tono de voz aumentaron.

b) Declaraciones planeadas o no planeadas. De los ocho indicios en los que se pudo comparar si había cambios en función de si las declaraciones habían sido planeadas previamente o no, solamente se observaron variaciones significativas producidas por esta variable en la latencia de respuesta. Cuando las declaraciones no se habían planeado, la latencia de respuesta era significativamente mayor al mentir que al decir la verdad. Cuando las declaraciones se habían planeado previamente, se observaba una tendencia opuesta, es decir, la latencia de respuesta era menor al mentir.

c) Duración de las declaraciones. Hubo tres indicios sobre los que se pudo establecer comparaciones en función de si la declaración era larga o corta: longitud de la respuesta⁶, latencia de respuesta y tono de voz. A diferencia de las declaraciones cortas, cuando las declaraciones debían prolongarse por más tiempo las respuestas falsas eran más cortas que las verdaderas, la latencia de respuesta era mayor y el tono de voz era más agudo.

d) Contenido de la declaración: transgresiones vs. no transgresiones. De los 12 indicios en los que se pudo comparar si había diferencias significativas en función de si el tema de la declaración era o no sobre una transgresión, en cuatro de ellos se observaron cambios significativos producidos por este moderador.

Tabla 7. *Tamaño del Efecto (d) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función de la Diferencia en el Contenido de la Declaración (DePaulo, Lindsay, et al., 2003).*

	No transgresión	Transgresión
Tasa del habla ^a	.01	.32*
Movimientos de pies o piernas	-.04	-.24*
Nerviosismo, tensión	.09	.51*
Pestañeos	.01	.38*
Mov. Nerviosos/inquietos	.24*	-.16

^aIndicio definido como número de palabras o sílabas por unidad de tiempo.

* $p < .05$.

De los cinco indicios que aparecen en la Tabla 7, cuatro (tasa del habla, movimientos de piernas o pies, tensión y pestañeos) no diferenciaron significativamente verdades de mentiras cuando el contenido de las declaraciones no tenía que ver con una transgresión, pero sí cuando se trataba de una transgresión. En concreto, a diferencia de las declaraciones sobre temas distintos de una transgresión, en las declaraciones sobre transgresiones cuando se mentía la tasa de habla era superior, se movían menos las piernas y los pies, se mostraba más tensión y se pestañeaba más que cuando se decía la verdad. Respecto a los movimientos nerviosos / inquietos, el patrón fue opuesto, es decir, esta conducta discriminó significativamente en los casos en los que el tema de la declaración no era sobre una transgresión (cuando la

⁶ Este indicio se refiere al tiempo que el declarante empleaba en contestar a cada pregunta del interrogador.

declaración era falsa, el declarante presentaba más movimientos inquietos), pero no si se trataba de una transgresión.

e) Medición objetiva de los indicios o impresiones subjetivas. De los cinco indicios no verbales en los que se pudo comparar si había diferencias en significatividad en función de si las conductas se habían medido de forma objetiva o subjetiva, solamente se observó un efecto significativo sobre la apariencia facial de bienestar. Cuando los indicios se habían estimado basándose en apreciaciones subjetivas, las personas que mentían se percibían significativamente con menos apariencia facial de bienestar que las que decían la verdad. No hubo diferencias entre verdades y mentiras cuando los indicios se habían medido de forma objetiva.

En definitiva, la comprobación de que, efectivamente, existe una serie de variables moderadoras que pueden afectar a la validez de los indicios no verbales de engaño, no hace sino añadir ambigüedad a una investigación ya de por sí complicada. Las conclusiones que se pueden extraer del trabajo de DePaulo, Lindsay, et al. (2003) en este sentido no son muy alentadoras. Si no existen indicios de engaño perfectos y solamente un pequeño grupo de ellos parecen tener una “imperfecta” utilidad predictiva, si, además, de los indicios de este pequeño grupo la mayor parte son de difícil observación o no han sido claramente definidos, y si, finalmente, la idiosincrasia de las situaciones en las que pueden aparecer determina que distintas variables moderadoras afecten a su validez como herramienta de discriminación, saber cuando alguien nos está mintiendo o diciendo la verdad a partir de la observación de la conducta no verbal ha de ser, sin duda, una tarea extremadamente difícil.

Meta-análisis de Sporer y Schwandt (2007)

En este trabajo los autores trataron de examinar la validez de determinados indicios no verbales para detectar mentiras. En total se seleccionaron 11 conductas visibles bajo la doble premisa de que fueran directamente observables y que estuvieran entre las que más comúnmente tanto profesionales como no profesionales tienen en cuenta a la hora de juzgar la veracidad. En total se analizaron los datos de 41 estudios. Uno de los objetivos

de Sporer y Schwandt (2007) era examinar el efecto de distintas variables moderadoras sobre los indicios de engaño que se iban a estudiar. Por lo tanto, se consideraron por separado las distintas condiciones o grupos de un mismo estudio que representaban dichas variables moderadoras. En consecuencia, se contó finalmente con 54 muestras o pruebas diferentes. Para los análisis de datos, los autores se basaron en la observación del tamaño del efecto de cada indicio al mentir. En concreto, para cada indicio se calculó el índice r (véase Sporer y Schwandt, 2007 para una explicación más detallada). Valores positivos de r indican un incremento de la conducta estudiada al mentir, y valores negativos una disminución.

En la Tabla 8 se presentan los resultados para cada uno de los 11 indicios a nivel general, es decir, sin tener en cuenta el posible efecto de variables moderadoras. Los datos están ponderados en función del tamaño de cada muestra. Como se puede observar, solamente tres de los 11 indicios analizados se mostraron significativamente asociados a la mentira: movimientos de asentimiento con la cabeza ($r = -.091$, $p = .014$), movimientos de las manos ($r = -.186$, $p = .001$) y movimientos de pies y piernas ($r = -.067$, $p = .029$). En los tres casos la relación es negativa, es decir, los tres tipos de conductas disminuyen al mentir. Curiosamente, ninguno de estos tres indicios discriminó significativamente en el meta-análisis de DePaulo, Lindsay, et al. (2003) a nivel global.

Debido a razones de tipo metodológico, Sporer y Schwandt (2007) sólo consideraron adecuado estudiar la influencia de las variables moderadoras sobre siete de los 11 indicios, en concreto sobre el contacto ocular, movimientos de la cabeza, asentir con la cabeza, sonrisas, adaptadores, ilustradores y movimientos de las manos. Resumimos a continuación sus hallazgos.

a) Contenido de la mentira. Se analizó si el contenido de las mentiras (sentimientos y hechos vs. sólo hechos), influía en la validez discriminativa de cada indicio. En la Tabla 9 presentamos los cinco indicios que diferenciaron significativamente verdades de mentiras en una de las condiciones pero no en la otra.

Tabla 8. Asociación Real de los Indicios Analizados por Sporer y Schwandt (2007) con el Engaño.

Indicio		K	N	r (media)	p
Pestañeos	Frecuencia con que ambos ojos se cierran rápidamente de forma simultánea.	11	769	.004	.456
Contacto ocular	Duración del tiempo que pasa el declarante mirando a los ojos al entrevistador.	15	791	-.007	.422
Evitación de la mirada	Duración o frecuencia del tiempo que pasa el declarante sin mirar al entrevistador.	5	303	.027	.324
Contacto ocular (combinado)	Combinación de contacto ocular y evitación de la mirada.	20	1094	-.011	.359
Movimientos de la cabeza	Frecuencia de movimientos arriba – abajo y derecha – izquierda de la cabeza.	7	362	.059	.131
<i>Asentir con la cabeza</i>	<i>Frecuencia de movimientos de delante hacia atrás de la cabeza sugiriendo afirmación claramente perceptibles.</i>	9	590	-.091	.014
Sonrisa	Duración o frecuencia de las sonrisas.	20	1134	-.032	.141
Adaptadores	Duración o frecuencia de movimientos en los que una mano toca otra mano o parte del cuerpo o la cara (ej. Rascarse, acariciarse).	27	1539	.018	.240
<i>Movimientos de las manos</i>	<i>Frecuencia de los movimientos de las manos sin que el brazo se mueva.</i>	5	308	-.186	.001
Ilustradores	Frecuencia de movimientos de las manos que acompañan al habla y acentúan, enfatizan o marcan el ritmo del discurso.	21	1355	.016	.287
<i>Movimientos de pies y piernas</i>	<i>Duración o frecuencia de movimientos de los pies y las piernas, tales como cruzar las piernas o mover los pies de forma rápida.</i>	15	799	-.067	.029
Cambios de postura	Frecuencia de movimientos hechos para cambiar la posición en la silla.	17	872	.011	.373

Nota. Se han puesto en cursiva los datos referidos a indicios que alcanzaron un valor significativo. K representa el número de comparaciones; N representa el número total de participantes.

Tabla 9. Tamaño del Efecto (r) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función del Contenido de la Declaración (Sporer y Schwandt, 2007).

	Sentimientos y hechos	Sólo hechos
Contacto ocular	-.127*	.032
Movimientos de la cabeza	.255**	.000
Asentir con la cabeza	-.223**	.055
Sonrisa	-.099*	-.003
Adaptadores	.104*	-.001

*p < .05; **p < .01

La Tabla 9 muestra que cuando las declaraciones implicaban hechos y sentimientos, los cinco indicios fueron capaces de discriminar significativamente entre verdades y mentiras. Sin embargo, ninguno de ellos discriminó significativamente cuando las declaraciones sólo implicaban hechos. En concreto, al mentir sobre sentimientos y hechos disminuye el contacto ocular, los asentimientos con la cabeza y las sonrisas en comparación con cuando se dice la verdad. Además se produce un aumento diferencial de los movimientos de la cabeza y los adaptadores.

b) Motivación. En condiciones de baja motivación (aquellas en las que se había ofrecido un pequeño incentivo monetario o la posibilidad de contratar al participante como parte del equipo), quienes mentían sonreían menos ($r = -.075$, $p = .037$), hacían más movimientos de cabeza ($r = .116$, $p < .05$), y asentían menos ($r = -.111$, $p < .05$) que quienes decían la verdad. En condiciones de alta motivación (aquellas en las que se había hecho especial hincapié en la importancia de saber mentir bien para el desarrollo profesional del individuo), solo los movimientos de las manos ($r = -.204$, $p < .01$) discriminaron significativamente. Quienes mentían movían menos las manos.

c) Preparación. Dado que ninguno de los estudios analizados había incluido periodos largos de preparación, sólo se consideraron preparación mínima y preparación media. En condiciones de mínima preparación, al mentir se producían menos movimientos de las manos ($r = -.226$, $p < .01$) y menos asentimientos ($r = -.158$, $p < .001$). En condiciones de preparación media, solamente los movimientos de la cabeza ($r = .146$, $p < .05$) discriminaron significativamente. Quienes mentían movían la cabeza más frecuentemente.

d) Finalmente, Sporer y Schwandt (2007) también encontraron que otro tipo de variables que no tenían que ver con las características de la situación en la que se mentía, sino más bien con la metodología experimental utilizada en cada estudio, afectaban a la significatividad de los indicios analizados. Por ejemplo, al comparar los estudios intra e inter-sujeto, se observó que en general los tamaños de los efectos eran mayores para los primeros. Del mismo

modo, se observaron diferencias en función de las distintas formas en que las variables se habían operacionalizado en los diversos estudios.

Meta-análisis de Sporer y Schwandt (2006)

Dado que en este apartado se han tratado conjuntamente indicios visibles y paralingüísticos, hemos considerado interesante incluir un meta-análisis que se ha centrado exclusivamente en estos últimos. Sporer y Schwandt (2006) sintetizaron los resultados de distintos estudios respecto a nueve indicios paralingüísticos: duración del mensaje, número de palabras, tasa del habla, latencia de respuesta, pausas vacías, pausas llenas, errores del habla, repeticiones y tono de voz. En total se analizaron un total de 41 estudios. Los análisis se realizaron ponderando o sin ponderar los datos en función de los tamaños muestrales⁷. En la Tabla 10 se presentan los resultados globales para cada uno de los nueve indicios. Tal como se aprecia en esta tabla, los resultados fueron diferentes en función de si se ponderaban o no los datos. La duración del mensaje, el tono de voz, la latencia de respuesta y los errores del habla discriminaron significativamente verdades de mentiras cuando se consideraron los datos de los diferentes estudios sin ponderar, es decir, dando el mismo valor a todos los estudios sin tener en cuenta las diferencias en tamaño muestral. Según estos resultados, los mentirosos emiten respuestas significativamente más cortas, hablan con un tono de voz más alto, tardan más en empezar a contestar y cometen más errores al hablar que quienes dicen la verdad.

Cuando se tuvo en cuenta el tamaño de cada una de las muestras consideradas, sólo el tono de voz y la latencia de la respuesta discriminaron significativamente. Cuando las personas mienten hablan con un tono de voz más agudo y tardan más en responder.

⁷ Para ponderar los datos en función del tamaño muestral se utilizaron dos métodos (véase Sporer y Schwandt, 2006). Dado que con ambas técnicas se observó una discriminación significativa para los mismos indicios, se mostrarán aquí los resultados siguiendo la metodología de Hedges y Olkin (1985).

Tabla 10. *Asociación Real de Cada Indicio con el Engaño (Sporer y Schwandt, 2006).*

Indicio	Definición	<i>Medias sin ponderar</i>		<i>Medias ponderadas</i>	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Duración del mensaje	Tiempo total que el declarante permanece hablando.	<i>-.057</i>	<i>.018</i>	<i>-.039</i>	<i>.074</i>
Número de palabras	Número total de palabras incluidas en la declaración.	<i>.012</i>	<i>.409</i>	<i>-.009</i>	<i>.433</i>
Tasa del habla	Número de palabras divididas por la duración del mensaje.	<i>.007</i>	<i>.413</i>	<i>.012</i>	<i>.354</i>
Pausas llenas	Alteraciones del discurso por la inclusión de vocalizaciones como “uh”, “er”, “um”, “ah”, etc.	<i>.026</i>	<i>.233</i>	<i>.040</i>	<i>.135</i>
Pausas vacías	Rupturas del discurso debidas a la inclusión de espacios de silencio.	<i>.017</i>	<i>.355</i>	<i>-.016</i>	<i>.365</i>
Tono de voz	Frecuencia de la voz, analizada mediante instrumentos electrónicos específicos.	<i>.132</i>	<i>.013</i>	<i>.101</i>	<i>.044</i>
Repeticiones	Palabras o frases duplicadas por razones distintas de errores del habla o preguntas del entrevistador.	<i>.112</i>	<i>.098</i>	<i>.084</i>	<i>.164</i>
Latencia de respuesta	Tiempo entre el final de la pregunta del entrevistador y el principio de la respuesta.	<i>.088</i>	<i>.004</i>	<i>.106</i>	<i>.001</i>
Errores del habla	Aparición de errores gramaticales (ej. Cambiar un singular por un plural).	<i>.059</i>	<i>.035</i>	<i>.041</i>	<i>.102</i>

Nota. Se han puesto en cursiva los datos de los indicios que discriminaron significativamente verdades de mentiras. El índice *r* representa la media del tamaño del efecto mostrado en los distintos estudios revisados.

Al igual que en el trabajo anterior, Sporer y Schwandt (2006), examinaron el efecto de diversas variables moderadoras sobre los indicios:

a) Contenido de la declaración. Se analizó si el diferente contenido de las narraciones (sentimientos y hechos vs. sólo hechos) influía en la validez discriminativa de cada indicio. En la Tabla 11 se muestra que el tono de voz y la latencia de respuesta discriminaron significativamente verdades de mentiras cuando la declaración implicaba tanto sentimientos como hechos (al mentir, aumentaba el tono de voz y la latencia de respuesta en comparación con cuando se decía la verdad), pero no cuando la declaración solamente implicaba hechos. Con la duración del mensaje los resultados fueron inversos, es decir, este indicio discriminó significativamente cuando las declaraciones eran

solamente sobre hechos (menor duración del mensaje al mentir que al decir la verdad) pero no cuando implicaban hechos y sentimientos.

Tabla 11. *Tamaño del Efecto (r) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función del Contenido de la Declaración (Sporer y Schwandt, 2006).*

	Sentimientos y hechos	Sólo hechos
Duración del mensaje	.014	-.075*
Tono de voz	.220**	-.051
Latencia de respuesta	.407**	.024

* $p < .05$; ** $p < .01$.

b) Preparación. En la Tabla 12 se observa que el tono de voz y la latencia de respuesta discriminaron significativamente verdades de mentiras cuando la declaración había sido mínimamente preparada (al mentir, aumentaba el tono de voz y la latencia de respuesta en comparación con cuando se decía la verdad), pero no cuando la declaración había sido medianamente preparada. Por su parte, la duración del mensaje mostró un patrón inverso, es decir, discriminó significativamente cuando las declaraciones habían sido medianamente preparadas (menor duración del mensaje al mentir que al decir la verdad) pero no cuando habían sido mínimamente preparadas.

Tabla 12. *Tamaño del Efecto (r) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función de la Preparación (Sporer y Schwandt, 2006).*

	Mínima preparación	Preparación media
Duración del mensaje	.006	-.068*
Tono de voz	.194**	-.061
Latencia de respuesta	.190**	-.044

* $p < .05$; ** $p < .01$.

c) Motivación. Se establecieron dos niveles de motivación para su comparación: bajo (condiciones en las que se había ofrecido un pequeño incentivo monetario o la posibilidad de contratar al participante como parte del equipo) y alto (condiciones en las que se había hecho especial hincapié en la importancia de saber mentir bien para el desarrollo profesional del individuo).

Tabla 13. *Tamaño del Efecto (r) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función de la Motivación (Sporer y Schwandt, 2006).*

	Baja motivación	Alta motivación
Duración del mensaje	-.016	-.102*
Tasa de habla	-.024	.102*
Tono de voz	.012	.529**
Latencia de respuesta	.049	.187**

* $p < .05$; ** $p < .01$.

Los cuatro indicios de la Tabla 13 fueron buenos discriminadores de verdades y mentiras cuando los declarantes estaban altamente motivados para ser creídos. En tales condiciones, al mentir las declaraciones eran más cortas, la tasa de habla mayor, el tono de voz más elevado y la latencia de respuesta mayor que al decir la verdad. Por otro lado, en condiciones de baja motivación, ninguno de estos indicios discriminó significativamente verdades de mentiras.

d) Mentiras solicitadas vs. no solicitadas por el investigador. En la Tabla 14 se muestra la duración del mensaje, la tasa de habla y las pausas llenas, variables que a nivel global no fueron buenos discriminadores de verdades y mentiras con los datos ponderados. Sí lo fueron, no obstante, en aquellos casos en los que el participante había mentido espontáneamente sin una petición explícita del investigador. Se podría pues concluir que ante la ausencia de petición, al mentir se muestran mensajes más cortos, una mayor tasa de habla y más pausas llenas que al decir la verdad. Cuando las mentiras no obedecían a ningún tipo de petición, ninguno de los tres indicios diferenció significativamente (Tabla 14).

Tabla 14. *Tamaño del Efecto (r) de la Relación Entre Mentir o Decir la Verdad y Distintos Indicios en Función de si las Mentiras Eran o no Solicitadas (Sporer y Schwandt, 2006).*

	Mentiras no solicitadas	Mentiras solicitadas
Duración del mensaje	-.228**	-.017
Tasa de habla	.209*	-.008
Pausas llenas	.325**	-.005

* $p < .05$; ** $p < .01$.

Finalmente Sporer y Schwandt (2006), al igual que en el estudio anterior, encontraron que otro tipo de variables que no tenían nada que ver con las características de la situación en la que se mentía, sino con la metodología utilizada en cada estudio, afectaban a la significatividad de los indicios

analizados. Por ejemplo, al comparar los estudios intra e inter-sujeto, se observó que en general los tamaños de los efectos eran mayores para los estudios inter-sujeto. Del mismo modo, como en el trabajo anterior, se observaron diferencias en función de la forma en que las variables habían sido operacionalizadas en los distintos estudios.

En definitiva, y como conclusión a este epígrafe, hay que decir que en los dos trabajos anteriores (Sporer y Schwandt, 2006, 2007) ha quedado patente que ciertas variables moderadoras hacen que determinados indicios discriminen entre verdades y mentiras en ciertas circunstancias pero no en otras. Esto hace aún más difícil encontrar respuesta a la pregunta “¿existen indicios no verbales realmente eficaces para detectar el engaño?” que encabezaba este apartado.

Los resultados de los tres meta-análisis presentados demuestran que no solamente no hay indicadores no verbales o para-verbales perfectos de engaño, sino que, además, la relación de las conductas analizadas con la mentira dista mucho de ser buena. Como se ha observado, además de ser pocos aquellos indicios que a nivel global son capaces de discriminar (probabilísticamente) verdades de mentiras, dicho potencial discriminador se ve afectado por numerosas variables moderadoras. De este modo, quizá habría que cambiar la pregunta inicial por otra más modesta, como por ejemplo “¿existen indicios no verbales realmente eficaces para detectar el engaño en una determinada situación?”, y clarificar el tipo de situación a la que nos refiriésemos. No obstante, aún teniendo en cuenta estos aspectos, la precisión de quienes tratan de detectar mentiras a través del análisis de la conducta no verbal podría estar por encima del azar si se utilizasen para ello los indicios adecuados, lo cual nos lleva a la segunda pregunta de este apartado, referente a los indicios que las personas usan para juzgar la veracidad.

B) ¿Cuáles son los indicios en los que las personas habitualmente basan sus juicios?

Antes de comenzar a responder esta pregunta, sería adecuado hacer una distinción general respecto a los indicios de mentira. Se ha denominado

indicios “reales” de engaño a aquellos que se han encontrado realmente asociados al acto de mentir, y a ellos nos hemos referido en el apartado anterior. Por otra parte estarían los indicios “percibidos” de engaño, que serían aquellos en los que se ha observado que realmente los individuos basan sus juicios de veracidad⁸.

Hay un acuerdo general entre los principales estudiosos de este campo en señalar que los indicios en los cuales los individuos más comúnmente suelen basar sus juicios de veracidad no coinciden en gran parte con los indicios realmente asociados al engaño (por ejemplo, Masip, 2005; Masip, Alonso y Herrero, 2006; Masip et al., 2002a; Vrij, 1998a, 2000; Vrij y Semin, 1996; Zuckerman et al., 1981).

Granhag y Vrij (2005) plantean que a la hora de estudiar cuáles son las conductas en las que las personas más frecuentemente nos basamos para concluir si alguien está mintiendo o diciendo la verdad, pueden utilizarse dos tipos de estrategias. Por un lado, puede pedirse directamente al individuo que indique los indicios que considera asociados a la mentira, por ejemplo a través de un cuestionario (estaríamos hablando de creencias sobre indicios de engaño). Este sería el caso de estudios como Granhag, Anderson, Strömwall y Hartwig (2004). En esta investigación, los autores pidieron a tres grupos de participantes (presos, trabajadores de prisiones y estudiantes) que señalaran en un cuestionario qué conductas de una lista consideraban asociadas a la mentira. En línea con trabajos anteriores, encontraron que los estudiantes y los trabajadores de prisiones mantenían creencias similares y erróneas sobre los indicios del engaño, mientras que las creencias de los presos se ajustaban más a la realidad.

⁸ La investigación ha mostrado que no está claro si los individuos son del todo conscientes de los indicios en los que realmente basan sus juicios de engaño (Vrij, 2000). Zuckerman, Koestner y Driver (1981, en Zuckerman et al., 1981) preguntaron explícitamente a los participantes sobre sus “creencias” en relación a distintos indicios de engaño, y luego correlacionaron éstas con los indicios en los que realmente se basaron los juicios. Los resultados mostraron que aunque existe una correlación significativa, ésta dista de ser perfecta. No obstante, dado que existe cierta coincidencia, en muchos estudios se habla únicamente de indicios percibidos de engaño y se engloba tanto a las creencias como a los indicios percibidos.

En este contexto, The Global Deception Research Team (2006) presenta un interesante trabajo en el que se examinan las creencias sobre indicios de engaño de muestras de individuos de 75 países distintos. En un primer estudio, un total de 2320 participantes (20 hombres y 20 mujeres por país) de 58 países distintos contestaron por escrito a la pregunta *¿How can you tell when people are lying?* traducida a la lengua específica de cada país. Se obtuvieron 11.157 respuestas a esta pregunta, las cuales fueron codificadas en 103 creencias distintas sobre indicios de engaño.

Los resultados mostraron que existen estereotipos compartidos por participantes de distintos países respecto a qué indicios se relacionan con el engaño. Las nueve conductas que los participantes creyeron más asociadas al engaño (citadas por al menos el 15% de los participantes de cada país) fueron, por este orden: evitación de la mirada, nerviosismo, incoherencias en las declaraciones, movimientos corporales, expresión facial, inconsistencias verbales, pausas llenas, enrojecimiento facial y pausas vacías. De todas estas creencias hubo una que sobresalió por encima de las demás: la evitación de la mirada, ya que el 63.66% del total de los participantes manifestaron confiar en este indicio a la hora de juzgar una declaración como falsa. El nerviosismo fue la siguiente creencia mayormente compartida (citada por un 28.15% de los participantes), aunque, como observamos, a una gran distancia de la primera.

El segundo de los estudios que constituyeron esta investigación consistió en presentar un cuestionario a otros participantes. Se les preguntaba si el engaño se relaciona con un conjunto específico de conductas entre las cuales se encontraban aquellas más mencionadas en el primer estudio. Las conductas fueron: contacto ocular, cambio de postura, tocarse / arreglarse alguna parte del cuerpo, longitud de la historia, tartamudear, duración de las pausas, nerviosismo, gestos con las manos, consistencia de la historia y seriedad. Además, para cada conducta concreta, se pedía que se especificase la dirección de la misma al mentir (aumenta, disminuye, permanece constante). En total se contó con muestras de 63 países distintos. Cada una de tales muestras estaba compuesta por 20 hombres y 20 mujeres.

En este segundo estudio se observa de nuevo que el estereotipo más compartido por los participantes se refiere al contacto ocular, ya que el 71.5% de ellos mostró la creencia de que éste disminuye al mentir. Otras creencias comunes fueron pensar que al mentir se cambia más de postura (65.2% de los participantes), creer que la persona que miente ejecuta más conductas referidas a tocarse / arreglarse alguna parte del cuerpo (64.8%) y creer que las historias falsas son más largas que las verdaderas (62.2% de los participantes).

En definitiva, esta investigación muestra que existen creencias erróneas culturalmente compartidas sobre los indicios no verbales de engaño. Ciertas conductas, como por ejemplo apartar la mirada, cambiar de postura, hacer movimientos corporales o tocarse / arreglarse alguna parte del cuerpo, se interpretan de forma general como indicios de mentira. Sin embargo, el meta-análisis de DePaulo, Lindsay, et al. (2003) ha mostrado que dichas conductas no se relacionan, a nivel global, con mentir o decir la verdad. De las nueve conductas más citadas en el primer estudio, solamente el nerviosismo (citado por un 28.15% de los participantes) se encontró significativamente asociado al engaño en el meta-análisis de DePaulo, Lindsay, et al. (2003). Más relevante todavía es que trabajos como el de Sporer y Schwandt (2007) hayan encontrado que en ocasiones las creencias populares (tales como que al mentir aumentan los movimientos de piernas, pies y manos) van en dirección opuesta a la relación real de dichas conductas con la mentira (estos autores observaron que al mentir disminuían los movimientos de piernas, pies y manos).

La segunda estrategia propuesta por Granhag y Vrij (2005) para saber cuáles son los indicios en los que las personas basan sus juicios de mentira consiste en presentar un conjunto de declaraciones a un grupo de individuos, y pedirles que emitan un juicio de veracidad sobre las mismas, indicando además en qué indicios basaron su juicio (seguiríamos hablando de creencias sobre los indicadores del engaño). También pueden establecerse relaciones entre el tipo de juicio emitido (verdad o mentira) y la presencia o ausencia de determinadas conductas en la declaración (ahora estaríamos hablando de indicios percibidos de engaño). Ejemplos de este tipo de estudios serían los de Bond, Omar, Mahmoud y Bonser (1990), Feeley y DeTurck (1995), Feeley y Young (2000) o Vrij y Winkel (1993).

Al revisar distintos estudios sobre indicios percibidos de engaño, Vrij (2000) ha concluido que los juicios de mentira se asocian con conductas como una elevación del tono de voz, aumento de errores y vacilaciones al hablar, mayores periodos de latencia, pausas más frecuentes y más largas durante el discurso, evitación de la mirada, sonrisas, más parpadeos y más movimientos (ilustradores, auto-manipulaciones, movimientos de manos, dedos, piernas, pies, tronco y cambios de postura). De todos estos indicios, solamente el tono de voz se ha encontrado realmente asociado con el engaño (DePaulo, Lindsay, et al., 2003; Sporer y Schwandt, 2006). Respecto a la latencia de respuesta, los resultados no son claros. Algunas revisiones como la del propio Vrij (2000), o el meta-análisis de DePaulo, Lindsay, et al. (2003), no han encontrado este indicio realmente asociado al engaño, mientras que en el meta-análisis de Sporer y Schwandt (2006) si se ha encontrado un aumento significativo al mentir. En la Tabla 15 se expone, a modo de resumen, una comparación entre los resultados de distintas revisiones y meta-análisis. Se observa que no hay una buena coincidencia entre indicios reales y percibidos de engaño.

En definitiva, los datos presentados implican una respuesta a la pregunta que abría este apartado; ¿cuáles son los indicios no verbales en los que las personas habitualmente basan sus juicios? La investigación muestra que aquellos individuos que, al evaluar la veracidad de una declaración, se decantan por un juicio de mentira, parecen basarse en conductas como un aumento de movimientos de las piernas, pies y brazos, cambios de postura, sonrisas, conductas referidas a tocarse o arreglarse alguna parte del cuerpo, aumento de pausas durante el discurso, y por encima de todo, desviar la mirada del interlocutor.

Otro aspecto, más importante si cabe, es que los distintos estudios presentados más arriba muestran que dichas conductas no tienen una relación real significativa con la verdad o la mentira. La siguiente y lógica pregunta que se plantea no puede sino hacer referencia a la precisión que se ha observado en distintos estudios a la hora de juzgar cuando alguien miente o dice la verdad.

Tabla 15. Comparación de Indicios de Engaño no Verbales Reales y Percibidos Según Distintas Revisiones y Meta-análisis.

Indicios	Indicios reales		Indicios percibidos					
	DePaulo, Lindsay, et al. (2003)	Sporer y Schwandt (2006, 2007) ^a	DePaulo et al. (1980)	Zuckerman et al. (1981)	DePaulo, Stone y Lassiter (1985)	Köhnken (1989)	Vrij (2000)	Sporer y Schwandt (2007)
Adaptadores	-	-					>	>
Apartar la mirada	-	-	>	>	>	>	>	>
Apretar los labios	>							
Asentir con la cabeza	-	<						
Bienestar facial aparente	<							
Cambios de postura	-	-	>	>		<		>
Cantidad de pausas	-	-					>	
Cooperatividad	<							
Dilatación pupilar	>							
Discrepancia / ambivalencia	>							
Duración de las pausas							>	
Elevación del mentón	>							
Errores del habla	-	-		>	>	>	>	
Ilustradores	<	-					>	>
Latencia de respuesta	-	>	>	>	>		>	
Movimientos inquietos	>							
Movimientos de las manos	-	<	>				>	>
Movimientos de piernas y pies	-	<					>	>
Movimientos del tronco	-						>	
Nerviosismo (apariencia global)	>		>					
Pestaños	-	-					>	>
Sonrisas	-	-	<	<	<	<	>	<
Tasa del habla	-	-		>	<	<	<	
Tensión vocal	>							
Tiempo hablando	<	-						
Titubeos / Vacilaciones	-			>	>	>	>	
Tono de voz	>	>		>	>	>	>	

Nota. El signo (>) indica que la frecuencia de aparición del indicio aumenta al mentir, el signo (<) que disminuye, y el signo (-) que no hay relación entre el indicio y la mentira.

^aLos resultados de Sporer y Schwandt (2006, 2007) sobre indicios reales que se presentan son los obtenidos con datos ponderados.

C ¿Cuál es la precisión en la detección de mentiras a partir de la observación de la conducta no verbal?

La gran mayoría de estudiosos dentro del campo de la detección del engaño coinciden en señalar que nuestra precisión a penas si está un poco por encima de lo meramente esperado por azar (Aamodt y Custer, 2006; Bond y DePaulo, 2006; Kalbfleisch, 1985). Presentaremos a continuación los resultados de dos meta-análisis realizados recientemente cuyo objetivo era estudiar la precisión en la detección del engaño: el de Aamodt y Custer (2006) y el de Bond y DePaulo (2006).

Aamodt y Custer (2006) llevaron a cabo un estudio meta-analítico para comprobar si había diferencias individuales en la habilidad de detectar el engaño. Se examinaron 108 estudios, que incluyeron más de 10.000 participantes agrupados como agentes de inteligencia, agentes de aduanas, agentes federales, agentes de libertad condicional, criminales, detectives, estudiantes, jueces, policías, profesores, psicólogos y trabajadores sociales. En la Tabla 16 se presenta la precisión media alcanzada por cada uno de estos grupos.

Tabla 16. *Porcentajes Medios de Precisión en Detección de Mentiras de los Distintos Grupos (Aamodt y Custer, 2006).*

Grupo	N Estudios	N Participantes	% Precisión
Agentes de inteligencia	1	34	64.12
Agentes de aduanas	3	123	55.30
Agentes federales	4	341	54.54
Agentes de libertad condicional	1	32	40.42
Criminales	1	52	65.40
Detectives	7	758	50.80
Estudiantes	156	11647	54.22
Jueces	2	194	59.01
Policías	12	655	55.30
Profesores	1	20	70.00
Psicólogos	4	508	61.56
Trabajadores sociales	1	20	66.25
TOTAL	193	14.379	54.50

Como se observa en la Tabla 16, algunos grupos (agentes de inteligencia, criminales, psicólogos y trabajadores sociales) parecen ser mejores detectores de mentiras que el resto. Aunque es posible que realmente se trate de colectivos especialmente habilidosos en la detección del engaño, hay que

tomar los datos con precaución, ya que en los cuatro casos se contó con muestras muy pequeñas y, excepto los psicólogos, cada grupo estaba representado por un sólo estudio. Cuando se dispuso de muestras mayores y más cantidad de estudios (detectives, estudiantes y policías), las tasas de precisión obtenidas fueron cercanas al nivel azar (50%). Este resultado también se observó al analizar conjuntamente todos los estudios (Tabla 16).

El meta-análisis de Bond y DePaulo (2006) incluyó muchos más estudios que el de Aamodt y Custer (2006). Estos autores sinterizaron los resultados de un total de 206 informes de investigación que ofrecían resultados de 177 muestras de emisores y 384 muestras de receptores. La precisión global observada fue de 53.46%, leve, pero significativamente por encima de lo meramente esperado por azar. Adicionalmente, en este estudio se observó una mayor precisión para la detección de verdades (61.34%) que para la detección de mentiras (47.55%). No obstante, esta diferencia no se debía a que los participantes realmente reconocieran mejor las verdades que las mentiras, sino a que, en general, se tiende más a hacer juicios de verdad que de mentira: el porcentaje de juicios de verdad emitidos fue del 55.23%, superior por lo tanto al total de juicios de mentira.

Los resultados de Bond y DePaulo (2006) respecto a la precisión en la detección del engaño coinciden en general con las tasas encontradas en revisiones y meta-análisis anteriores, tal como se muestra en la Tabla 17. No obstante, el trabajo de Bond y DePaulo (2006) es más exhaustivo y reciente, por lo que merece mayor consideración.

Tabla 17. *Resultados de Meta-análisis y Revisiones Sobre Precisión en la Detección del Engaño a Partir de Indicios no Verbales.*

Estudios	Precisión (%)
Kraut (1980)	57.0
Kalbfleisch (1985)	57.5
Vrij (2000)	56.6
Bond y DePaulo (2006)	53.5

Los porcentajes presentados en la Tabla 17 no deben sorprendernos teniendo en cuenta que, tal y como se ha observado anteriormente, no solo no hay indicios altamente fiables de engaño, sino que, además, a nivel general,

tendemos a basar nuestros juicios de veracidad en conductas que la investigación científica ha mostrado que carecen de valor diagnóstico.

Conclusiones

Tanto los primeros estudios sobre la detección del engaño a partir de indicios conductuales no verbales, como la investigación más reciente, coinciden al ofrecer una serie de conclusiones generales.

a) En primer lugar, y quizá ésta sea la conclusión más interesante, no se ha encontrado ningún indicio o conducta que de forma inequívoca y en todos los casos indique que una persona está mintiendo o diciendo la verdad, esto hace que la detección del engaño desde esta perspectiva no sea una tarea fácil.

b) Sí se ha observado que existen algunos indicios que, aunque no sean fiables al cien por cien, aparecen con una frecuencia significativamente mayor al mentir que al decir la verdad. Los meta-análisis más recientes indican que un aumento en el tono de voz, una conducta (a nivel general) nerviosa, el aumento de la dilatación pupilar o de la tensión vocal, un aumento de la latencia de respuesta y la disminución de asentimientos con la cabeza y movimientos de pies, piernas y manos, se asocian realmente a la presencia del engaño (DePaulo, Lindsay, et al., 2003; Sporer y Schwandt, 2006, 2007). Los problemas que presentan algunas de estas conductas son su difícil observación directa o una falta de clarificación conceptual.

c) La validez discriminativa general de algunos de estos indicios se puede alterar significativamente como resultado de distintas variables moderadoras. Éstas hacen que las características idiosincrásicas de la situación pueda determinar el poder de un indicio para distinguir verdades de mentiras.

d) La investigación también ha encontrado que cuando tratamos de detectar el engaño no solemos utilizar los indicios adecuados, sino otros como la dirección de la mirada, los cambios de postura, conductas relacionadas con arreglarse o tocarse alguna parte del cuerpo (como arreglarse el pelo) y otros ya señalados. Se ha demostrado que estas acciones no están significativamente relacionadas con el hecho de mentir.

e) Todo lo anterior lleva a la quinta conclusión general, que se refiere a la falta de precisión característica de nuestros juicios de veracidad. Así pues, nuestra precisión global se sitúa en el 54%, apenas por encima del 50% esperado por azar.

f) Cuando tenemos que juzgar la veracidad de lo que nos dicen, solemos mostrar una mayor tendencia a hacer juicios de verdad que de mentira (sesgo hacia la verdad). Ello repercute en una mayor precisión en la detección de verdades que de mentiras (Levine, Park y McCornack, 1999).

Ante este panorama realmente poco esperanzador, algunos investigadores han mostrado su escepticismo sobre la investigación con indicios conductuales no verbales. No obstante, tal como Masip, Alonso y Herrero (2006) plantean, la conducta no verbal como indicio para evaluar la credibilidad de un individuo podría ser de utilidad en determinadas situaciones en las que otro tipo de información escasea, es necesario tomar una decisión a corto plazo y existen razones para sospechar que el individuo está mintiendo.

Por otro lado, tal como indicaban Vrij y Winkel (1993), una primera y provisional evaluación de la credibilidad a partir de la conducta no verbal podría también ayudar a los agentes de la ley a decidir si se prolongará la entrevista policial o si se realizarán otras posteriores en aquellos casos en los que no se dispone de evidencia adicional en esos momentos. Así pues, consideramos que es necesario seguir trabajando en esta aproximación.

La investigación con configuraciones de indicios: una alternativa de futuro

Uno de los problemas que condiciona toda la investigación con indicios conductuales no verbales es sin duda su limitado poder diagnóstico. No obstante, investigaciones recientes muestran que utilizar como predictor un determinado conjunto o configuración de indicios, en vez de un solo indicio o varios considerados individualmente, podría ser un sistema más eficaz para detectar mentiras. Esto se ha observado especialmente en aquellos estudios que han trabajado con redes neuronales artificiales.

Las redes neuronales artificiales (RNA) son sistemas de procesamiento de la información cuyo funcionamiento está inspirado en las redes neuronales biológicas. De forma breve, la manera en que las RNA operan se establece en tres fases: (a) entrada de datos, (b) procesamiento y (c) salida o respuesta. Se trata de sistemas que constan de un cierto número de elementos de procesamiento simples (los nodos o neuronas) que se organizan por capas, de forma que cada neurona se conecta con otras a través de enlaces de comunicación, cada uno de los cuales tiene asociado un peso que se va modificando con la entrada en el sistema de nueva información relevante. De esta forma, la RNA es capaz de aprender relaciones entre varios estímulos *input* presentados simultáneamente a través de múltiples ensayos, y puede ser programada para ofrecer una respuesta específica *output* al reconocer una determinada configuración de estímulos entrantes (véase Montaña y Jiménez, 2001, 2002; Pitarque, Roy y Ruiz, 1998; para una explicación más detallada del funcionamiento de las RNA).

Desde el campo de estudio de la detección del engaño, distintos investigadores han aplicado las ventajas de la capacidad de procesamiento de las RNA a la detección del engaño. Por ejemplo Rothwell, Bandar, O'Shea, y McLean (2006) desarrollaron el *Silent Talker*, un sistema automatizado no invasivo capaz de recoger la conducta no verbal de un individuo mientras emite una declaración y analizar la presencia de configuraciones de indicios simultáneamente para, finalmente, tomar una decisión sobre la veracidad de la declaración.

La forma de proceder fue la siguiente: En primer lugar, los investigadores introdujeron en el programa los indicios no verbales que debían ser considerados por la red. Seguidamente presentaron al sistema, basado en una RNA, un conjunto de declaraciones verdaderas y falsas, introduciendo con cada declaración información sobre la veracidad real de dichas declaraciones. A través de sucesivos ensayos, el *Silent Talker* fue capaz de identificar un perfil (más preciso con cada nuevo ensayo) para las declaraciones verdaderas y otro para las falsas a partir del análisis de la presencia de la configuración de indicios que de forma más precisa se correspondía con cada tipo de declaración (véase Rothwell et al., 2006 para una explicación más detallada de

todo el proceso). Finalmente, los investigadores decidieron probar el sistema presentándole un conjunto de nuevas declaraciones (diferentes a las empleadas en la fase de entrenamiento). La precisión global obtenida estuvo por encima de los niveles del azar (80% cuando se había trabajado con declaraciones no planificadas, tanto en la fase de entrenamiento como en la de prueba; 61% cuando se había trabajado con declaraciones planificadas, tanto en la fase de entrenamiento como en la de prueba).

Aunque la aplicación de las RNA a la detección del engaño abre nuevos caminos de estudio en este campo, especialmente al mostrar cómo es más eficaz trabajar con configuraciones de conductas en lugar de indicios aislados, quedan aún importantes aspectos por clarificar, como por ejemplo determinar qué configuraciones de indicios son más eficaces con cada tipo de mentira o situación. Además, hay que tener en cuenta que, en el estudio de Rothwell et al. (2006), fue un programa informático a través de una RNA quien ejecutó la tarea de detección. Es necesaria más investigación para observar si los seres humanos podrían trabajar y tomar decisiones a partir de configuraciones de indicios. En este sentido, existe la posibilidad de que en un futuro cambie el rol del ser humano como “detector”, pasando más bien a ser un “gestor” de sistemas automatizados de detección. Junto a la investigación desarrollada por Rothwell et al. (2006), otros equipos están trabajando en esta línea de investigación a través del desarrollo de herramientas automatizadas de detección del engaño. Por ejemplo, Jensen, Meservy, Kruse, Burgoon y Nunamaker (2005) han obtenido tasas de entre el 70% y el 100% de precisión en la distinción de verdades y mentiras utilizando sistemas automatizados para detectar el engaño a partir de grabaciones en vídeo.

APROXIMACIÓN BASADA EN INDICADORES VERBALES

Desde esta aproximación se pretende detectar el engaño analizando el contenido verbal del discurso (Masip y Garrido, 1999, 2001a; Masip, Garrido y Herrero, 2003b). A nivel metodológico, una importante diferencia entre la aproximación basada en indicios conductuales no verbales que acabamos de describir y ésta es que en este tipo de investigación se han elaborado instrumentos o protocolos de aplicación sistematizada para el análisis del

discurso (Masip y Alonso, 2006). Uno de estos protocolos, quizá el que más se ha empleado, sobre todo en casos de abusos sexuales a niños, es la SVA (*Statement Validity Assessment*; en español: Evaluación de la Validez de la Declaración) (Garrido y Masip, 2001; Raskin y Esplin, 1991a; Ruby y Brigham, 1997; Steller, 1989; Steller y Koehnken, 1990/1994; Vrij, 2000, 2005) que describimos a continuación.

La SVA

La SVA se desarrolló en Alemania como una herramienta para determinar la credibilidad del testimonio de los niños en juicios sobre abusos sexuales (véanse Garrido y Masip, 2001; Granhag y Vrij, 2005; Steller y Koehnken, 1990/1994; Undeutsch, 1984, 1989; Vrij, 2000). Actualmente, la SVA se estructura en tres fases. En primer lugar, se realiza una entrevista semiestructurada al niño que alega haber sido sexualmente abusado. En segundo lugar, se aplica el CBCA (*Criteria-Based Content Analysis*; en español: Análisis de Contenido Basado en Criterios) sobre una transcripción literal de la entrevista. El CBCA valora sistemáticamente el contenido y calidad de la declaración al cotejar la información obtenida en la entrevista con una serie de “criterios de realidad”. Finalmente, las conclusiones obtenidas con el CBCA se cotejan con una “lista de validez” en la que se incluyen otros aspectos e información relevantes para el caso que se está juzgando (Horowitz, 1991; Masip y Garrido, 2001a, 2007; Raskin y Esplin, 1991a).

Con el fin de no alargar en exceso este apartado, en esta Tesis no se hará una exposición detallada de los orígenes la SVA, ni de cada uno de sus componentes, ni de protocolos previos como el SRA (Análisis de la Realidad de la Declaración) que la precedieron, y a partir de los cuales se elaboró. No obstante, el lector interesado puede consultar diversas fuentes (por ejemplo, Garrido y Masip, 2001a; Manzanero, 2001; Masip, 2002; Masip et al., 2003b; Raskin y Esplin, 1991a; Ruby y Brigham, 1997; Undeutsch, 1984, 1989). Sí es importante indicar que en la base teórica de la SVA, y de su fase nuclear, el CBCA, está la hipótesis que planteó el psicólogo alemán Udo Undeutsch, según la cual una declaración que se derive del recuerdo de una experiencia realmente vivida se diferenciará respecto al contenido y calidad de una

declaración basada en experiencias no vividas, es decir inventada o fantaseada (por ejemplo, Undeutsch, 1982). Este supuesto es conocido en la investigación sobre detección del engaño como “Hipótesis de Undeutsch”. Hecho este breve apunte histórico, seguidamente se describirá cada una de las tres partes de la SVA.

A) La entrevista semiestructurada

Raskin y Esplin (1991a) plantean que el objetivo de la entrevista es obtener información no sesgada a través de una narración libre. Posteriormente se valora en qué medida esta narración se ajusta a acontecimientos realmente vividos. En sentido estricto, pues, tal como han indicado algunos investigadores (por ejemplo Köhnken, Schimossek, Aschherman y Höfer, 1995), la entrevista semiestructurada, al igual que el resto de componentes de la SVA, no es un instrumento para evaluar la veracidad de una declaración. Se trata de una herramienta que pretende verificar la hipótesis previa de que las alegaciones del niño son ciertas. Solamente si no es posible dicha verificación, se acude a un conjunto de otras hipótesis alternativas que podrían explicar la declaración del niño. Esta es una idea que conviene tener presente para comprender la lógica de todo el proceso. Las hipótesis alternativas más frecuentemente consideradas son: (a) las alegaciones son básicamente ciertas, pero el niño ha sustituido al perpetrador por una persona diferente; (b) las alegaciones fundamentales son ciertas, pero el niño ha inventado o ha sido influenciado para añadir otras alegaciones falsas; (c) el niño ha sido influenciado o presionado para hacer una alegación completamente falsa en provecho de otra persona; (d) el niño ha hecho una alegación falsa por motivos personales de venganza, para obtener alguna recompensa o para ayudar a alguien; y (e) el niño ha inventado totalmente las alegaciones, posiblemente por problemas psicológicos (Raskin y Esplin, 1991a; véase también Garrido y Masip, 2001; Köhnken, 2004; Masip y Garrido, 2007).

Muchos estudiosos han hecho especial hincapié en que la información obtenida con la entrevista debe ser de calidad, es decir, libre de sesgos introducidos por el propio entrevistador sobre el recuerdo del niño (Garrido y Masip, 2001; Granhag y Vrij, 2005; Masip y Alonso, 2006; Raskin y Esplin,

1991a). Además, es muy importante conocer los posibles efectos de la memoria sobre el recuerdo de lo sucedido, especialmente si ha pasado algún tiempo desde que sucedió el supuesto episodio de abuso hasta que se hace la entrevista (Manzanero, 2001). Por todo ello, es fácil concluir que la entrevista debe realizarla un profesional cualificado, conocedor de las técnicas adecuadas para este tipo de situaciones.

La entrevista se estructura en tres partes:

a) La toma de contacto y establecimiento del *rapport*. Supone un primer acercamiento al niño, una “ruptura del hielo” por parte del profesional para ganarse, en la medida de lo posible, la confianza del menor.

b) El cuerpo de la entrevista, que es el momento en que el niño debe hacer su declaración en un estilo de narración libre. Solamente al final de la misma (si es necesario) el entrevistador plantea preguntas clarificadoras. En este sentido, Raskin y Esplin (1991a) hacen una buena clasificación de los tipos de preguntas a emplear en la entrevista y cuándo utilizar cada una. En primer lugar, el entrevistador debe comenzar invitando al niño a hacer una narración libre de los hechos. Si éste no menciona espontáneamente la situación, el entrevistador tratará de introducir el tema de una forma general no sesgada. En segundo lugar, pueden introducirse preguntas focalizadas una vez que el niño ha terminado su narración libre de lo acontecido. Éstas son preguntas abiertas que se plantean para obtener información o detalles necesarios para completar descripciones de hechos referidos durante la declaración. A continuación, si la descripción de los hechos no ha quedado suficientemente clarificada con las preguntas focalizadas, se puede pasar a las preguntas directas, cuyo objetivo es aclarar detalles específicos, ambigüedades o inconsistencias. Finalmente pueden formularse preguntas probatorias⁹, que son las menos abiertas de todas. No deben emplearse en aquellos casos en los que la hipótesis principal (las alegaciones del niño son ciertas) parece recibir un fuerte apoyo a la vez que las hipótesis alternativas parecen poco probables. Pueden utilizarse estas

⁹ Raskin y Esplin (1991a) denominan a este tipo de preguntas *Probing questions*. El término *probing* no tiene una traducción en español que refleje exactamente el significado de esta palabra en este contexto. Otros trabajos en lengua española (por ejemplo, Garrido y Masip, 2001) se refieren a este tipo de preguntas como “Preguntas para el contraste de hipótesis”.

preguntas, no obstante, si el niño describe algo que parece improbable, si parecen existir contradicciones en la declaración del niño o entre la declaración del niño y la de otras personas o si el entrevistador sospecha que el niño ha sido coaccionado por otras personas para declarar.

c) Finalización de la entrevista. Es el momento en el que el entrevistador considera que ha obtenido suficiente información y la entrevista puede finalizar.

Aquellas personas interesadas en una descripción más detallada de cómo desarrollar la entrevista y sus distintas fases, pueden consultar distintas fuentes (por ejemplo, Garrido y Masip, 2001; Masip, 2002; Masip y Garrido, 2007; Raskin y Esplin, 1991a) El lector interesado puede consultar adicionalmente el trabajo de Steller y Boychuk (1992) sobre los niños como testigos en casos de abusos sexuales, en el que incluyen como ejemplo la transcripción de una entrevista de un caso real de abuso sexual en el que se observa cómo se van introduciendo los distintos tipos de preguntas.

B) El CBCA (Análisis de Contenido Basado en Criterios)

Una vez finalizada la entrevista, se debe proceder a una transcripción literal de la misma, ya que será sobre esta transcripción sobre la que se apliquen los criterios del CBCA. En las primeras versiones de lo que hoy es el CBCA hubo modificaciones de los criterios que lo componían (véase por ejemplo Manzanero, 2001) hasta llegar a la versión más comúnmente aceptada de Steller y Köhnken (1990/1994). Estos autores sistematizaron una serie de trabajos anteriores, proponiendo un total de 19 criterios organizados en 5 categorías: (a) características generales; (b) contenidos específicos; (c) peculiaridades del contenido; (d) contenidos referentes a la motivación; y (e) elementos específicos del delito.

Antes de describir más detalladamente el CBCA y su aplicación es necesario considerar una idea básica que se establece como eje de todo el proceso: el concepto de “criterio de realidad” hace referencia a una característica verbal que puede aparecer en mayor o menor grado en una declaración. Cuanto más fuertemente presentes estén los criterios de realidad en una narración, mayor es la probabilidad de que ésta se corresponda con un acontecimiento

realmente vivido (Masip y Alonso, 2006; Masip y Garrido, 2001a, 2007; Raskin y Esplin, 1991a; Steller y Boychuk, 1992; Vrij, 2005). Se presentan a continuación los 19 criterios definidos y clasificados en las cinco categorías tal y como aparecen en Masip y Alonso (2006)¹⁰.

Características generales

Se refieren a la declaración tomada en su totalidad

1. Estructura lógica

Los diversos detalles describen idéntico curso de sucesos, la declaración en su conjunto es coherente y lógica, sus partes “encajan, la declaración tiene sentido.

2. Elaboración no estructurada

La declaración está algo desorganizada, es poco lineal, hay cambios de enfoque, los elementos relevantes están dispersos y la secuencia de sucesos no está en orden cronológico.

3. Cantidad de detalles

Descripciones de lugares, personas, objetos; concreción del momento temporal de los hechos, detalladas explicaciones de acontecimientos, etc. Las repeticiones del mismo detalle no deben contabilizarse de nuevo.

Contenidos específicos

Alguien que inventara la declaración no sería capaz de incluir en la misma estos contenidos, ya que ello superaría sus capacidades cognitivas

4. Engranaje contextual

El acontecimiento está inserto en un contexto espacio-temporal rico y complejo.

¹⁰ Para una descripción más detallada de cada criterio, el lector puede acudir a otras fuentes (por ejemplo, Masip y Garrido, 2007).

5. Descripción de interacciones

Acciones y reacciones mutuas entre el testigo y el agresor. Debe haber un mínimo de tres elementos.

6. Reproducción de conversaciones

Reproducción de un diálogo en que hay una réplica virtual de las palabras de al menos una persona, como en un *role playing*.

7. Complicaciones inesperadas

Dificultades que impiden la natural terminación del suceso.

Peculiaridades del contenido

Alguien que inventara la declaración no sería capaz de incluir en la misma estos contenidos, ya que ello superaría sus capacidades cognitivas.

8. Detalles inusuales

Detalles extraños, algo inesperados y sorprendentes, pero no irreales.

9. Detalles superfluos

Detalles periféricos que no contribuyen directamente a la alegación (por ejemplo, el color de la pared).

10. Incomprensión de detalles

Acciones o detalles que el testigo describe con exactitud pero que no comprende y que sí entiende el entrevistador.

11. Asociaciones externas relacionadas

Referencias a otros acontecimientos de tono sexual que tuvieran lugar fuera del incidente específico (por ejemplo, previas alusiones verbales al sexo por parte del agresor dirigidas a la víctima).

12. Alusiones al propio estado mental

Descripción de los propios sentimientos y pensamientos durante el transcurso del incidente.

13. Atribución del estado mental del autor del delito

Emociones, pensamientos y motivos que el narrador atribuye al supuesto autor del delito.

Contenidos referentes a la motivación

Alguien que de forma deliberada ofreciera un testimonio falso no introduciría estos contenidos, porque le restarían credibilidad.

14. Correcciones espontáneas

Corregirse uno mismo de forma espontánea

15. Admisión de falta de memoria

Admitir no acordarse de determinados detalles del acontecimiento.

16. Plantear dudas

Dudar de que otros crean la alegación

17. Autodesaprobación

Censurar la propia conducta.

18. Perdonar al autor del delito

No culpar o perdonar al agresor.

Elementos específicos del delito

19. Detalles característicos del delito

La descripción de los hechos se aleja de las falsas creencias legales sobre este tipo de delito y se acerca a cómo sucede realmente.

Hay que tener en cuenta algunos aspectos a la hora de aplicar los criterios a la declaración: (a) solamente deben considerarse aquellos criterios que aparezcan en la declaración cuando ésta se refiera al incidente sexual o a aspectos relacionados con éste; (b) las partes de la declaración que se repiten

solamente se puntúan una vez; (c) alguna información ofrecida puede puntuar en más de un criterio¹¹ (Manzanero, 2001; Steller y Boychuk, 1992). Garrido y Masip (2001) indican, además, que las respuestas a preguntas directas que no vayan más allá de lo que se dice en dicha pregunta no deben codificarse.

Para la puntuación de los criterios en la declaración se han planteado distintas propuestas, sin que ninguna haya llegado a constituirse como regla definitiva. Por ejemplo, Steller y Köhnken (1990/1994) propusieron dos posibilidades: bien evaluar cada criterio simplemente como presente o ausente o bien puntuar cada criterio en función de la fuerza o grado con que se cumple en la declaración. En otros trabajos de corte teórico (Manzanero, 2001; Steller y Boychuk, 1992; Vrij, 2000, 2005; Vrij y Akehurst, 1998) se alude a una valoración de la presencia de criterios en tres posibilidades: ausente, presente y fuertemente presente. La valoración de la declaración obtenida con el CBCA, independientemente del método de puntuación utilizado, no debe suponer en ningún caso una decisión definitiva sobre la veracidad de la declaración. Para tomar una decisión final al respecto hay que aplicar una herramienta más, la lista de validez.

C) La lista de validez

Los resultados del CBCA deben evaluarse conjuntamente con otro tipo de información o evidencia sobre el hecho procedente de fuentes distintas de la declaración. Esta información se explicita en la lista de validez.

Al igual que el CBCA, la lista de validez ha experimentado diversas modificaciones. En el Anexo A se presenta un esquema de la misma tal como aparece en Masip y Garrido (2007), que plantean una estructura similar a la que se propone en trabajos anteriores (por ejemplo, Raskin y Esplin, 1991a).

En este esquema hay dos categorías: I) Factores relacionados con la declaración, que se refieren a aspectos relacionados con la entrevista que podrían afectar a la validez de la declaración obtenida; y II) cuestiones

¹¹ Esto se da, por ejemplo, si mientras se describe una interacción se reproduce literalmente una conversación: “me acariciaba la espalda mientras me decía «tu me quieres mucho ¿verdad?» y yo le dije «sí»”.

investigativas, que abarcan una serie de “pruebas o información” que van más allá de la mera entrevista (Masip y Garrido, 2007).

Raskin y Esplin (1991a) explican la lista de validez transformando cada uno de los trece elementos en una pregunta cuya respuesta puede modificar o dar apoyo a los resultados del CBCA. Respecto a las *características psicológicas* habría que plantearse, según estos autores, los siguientes interrogantes: Primero, ¿existe alguna indicación de que ciertas limitaciones cognitivas o emocionales o sentimientos de incomodidad durante la entrevista puedan haber afectado a la calidad de la declaración? Si es así, habría que especificar qué limitaciones han afectado a la entrevista y de qué manera. Segundo, ¿cómo es el uso del lenguaje del niño? ¿Muestra un conocimiento normal para su edad, o se observa por sus expresiones que su conocimiento supera lo que podría haber aprendido con el incidente? Según Masip y Garrido (2007), si es así podría sospecharse que algún adulto lo ha instruido para declarar. Tercero, el afecto y las emociones mostradas por el niño durante la entrevista ¿son las que cabría esperar en este tipo de situaciones? Masip y Garrido indican no obstante que este es un criterio ambiguo en la medida en que no está claro cuáles han de ser las emociones “esperables”. Cuarto ¿en qué medida el niño se ha mostrado susceptible a las preguntas que le ha hecho el entrevistador? Es posible que el niño buscase pistas o indicios sobre qué decir al entrevistador.

Sobre las *características de la entrevista*, Raskin y Esplin (1991a) se plantean dos preguntas: Primera, ¿se han seguido las pautas de la SVA para realizar la entrevista semiestructurada? Si la respuesta a esta pregunta es negativa, cualquier resultado obtenido con el CBCA carece de valor. Segunda, ¿se han utilizado preguntas sugestivas, o algún tipo de presión o coerción sobre el niño? Esta pregunta viene a complementar la anterior. Recordemos que en la entrevista semiestructurada no se excluía la posibilidad de utilizar preguntas focalizadas o directas, pero siempre en el momento y de la forma adecuada. En ningún caso se debe admitir la presencia de preguntas sugerentes o capciosas.

Respecto a la *motivación para hacer la declaración*, hay que plantearse varios interrogantes: Primero, ¿cuáles han sido las circunstancias y el contexto en el que el niño ha hecho la alegación original? Se trata de examinar si en el contexto en que se realizó la alegación por primera vez hay elementos que puedan haber determinado de alguna forma la declaración del niño. Segundo, en relación con lo anterior, se puede plantear si el niño pudo haber recibido alguna presión de algún adulto para hacer la alegación. Finalmente, tal como plantean Raskin y Esplin (1991b), cabe preguntarse si la relación del niño con el acusado, u otras variables relacionadas, sugieren motivos para pensar que el menor pueda estar haciendo una falsa acusación.

Los cuatro últimos factores de la lista de validez componen una categoría que Masip y Garrido (2007) han denominado *cuestiones investigativas*. Hace referencia a aspectos que van más allá de la entrevista. Hay cuatro interrogantes que se pueden plantear:

a) ¿Son realistas los hechos descritos? Con esta pregunta se trata de comprobar si los principales elementos de la alegación pudieron realmente ocurrir o si desafían las leyes de la naturaleza, lo que los haría implausibles.

b) ¿Existe alguna otra declaración (bien hecha por el propio niño o bien ofrecida por algún otro testigo) que contradiga o sea inconsistente con los elementos principales de la alegación actual? Se trata de comprobar si los elementos esenciales y definitorios del episodio de agresión se mantienen constantes en otras posibles declaraciones. A este respecto, hay que centrarse en los elementos esenciales, no en detalles periféricos.

c) ¿Existe algún tipo de evidencia física fiable que vaya en contra de los elementos principales de la alegación? Para responder a esta pregunta sería necesario contar con la experiencia de otros profesionales, como médicos o especialistas, que pudieran certificar la existencia de pruebas (por ejemplo, análisis de ADN) o realizar exámenes fisiológicos al niño. Se trata de comprobar si la evidencia física disponible ofrece apoyo a la alegación.

d) Finalmente, ¿la descripción del hecho acontecido contiene los detalles y características generales normales de este tipo de delito? ¿Tiene la alegación

elementos importantes contrarios a las falsas creencias que individuos “no profesionales” mantienen dentro del campo de los abusos sexuales, pero que expertos en el tema han encontrado que son propios de este tipo de delitos? La presencia de este tipo de elementos sería un signo de verdad, ya que un adulto que hubiera presionado a un niño para hacer una alegación de este tipo, lo haría siguiendo sus esquemas o creencias, y no describiendo las pautas reales que en muchas ocasiones siguen este tipo de sucesos y que a menudo contradicen los estereotipos de los no expertos.

Todos estos interrogantes componen la lista de validez (véase, por ejemplo, Garrido y Masip, 2001; Masip y Garrido, 2007, para una descripción más detallada de cada uno de los trece factores), en la que nos hemos detenido especialmente para reivindicar su importancia, dado que es una parte de la SVA que muchas veces no es suficientemente tenida en cuenta.

En este apartado se han repasado brevemente cada una de las tres partes de la SVA. A continuación revisaremos también los resultados de distintos trabajos que han buscado determinar si esta herramienta es adecuada para ser empleada para juzgar la veracidad.

¿Es la SVA útil para distinguir verdades de mentiras? Limitaciones

Aunque, como ya hemos dicho, la SVA no pretende ser un instrumento para diferenciar verdades de mentiras, en muchos casos se ha utilizado como tal. Además, se ha empleado no solamente con niños y casos de abusos sexuales, sino también con adultos y otro tipo de delitos. La evidencia de la SVA es actualmente aceptada en los juicios en algunos tribunales de Estados Unidos (Ruby y Brigham, 1997) y diversos países europeos, como Alemania, Suiza, Austria, Holanda, Suecia o el Reino Unido (Köhnken, 2004). También en España.

Un primer problema a la hora de hacer una valoración de la utilidad de la SVA es que la mayoría de los estudios se han centrado en evaluar la utilidad de alguna de sus partes de forma independiente, sobre todo del CBCA, pero no de la SVA como un todo integrado. Esto supone juzgar “el todo” por “la parte”, lo cual es un grave error, ya que los resultados del CBCA por si solos no deben

suponer en ningún caso un diagnóstico definitivo, sino que deben siempre cotejarse con la evidencia de la lista de validez. Además, a la hora de comparar diversos estudios, se observa que no todos han trabajado con los 19 criterios del CBCA, sino que muchos han utilizado un subconjunto diferente de ellos, con lo que se hace difícil cotejar los resultados finales.

La entrevista semiestructurada y la lista de validez: algunas limitaciones

Uno de los problemas de la entrevista semiestructurada es que en muchos casos, especialmente en estudios de campo, simplemente no se ha aplicado, y frecuentemente las declaraciones se han obtenido a partir de entrevistas policiales que no siempre han seguido las directrices indicadas para este tipo de entrevistas. En otros casos, tal como se indica en Masip y Garrido (2007), aunque las entrevistas hayan sido hechas por profesionales, no se han seguido las recomendaciones establecidas sobre el modo de realizarlas (véase Masip y Alonso, 2006). A este respecto, Garrido y Masip (2001) observan que predominan las preguntas directas (aquellas que centran la atención del niño en aspectos del suceso que ya ha mencionado) o directivas (las que centran la atención del niño en aspectos del suceso que no ha mencionado) sobre las abiertas. Además, no siempre se han tenido en cuenta los posibles efectos del tiempo sobre el recuerdo, así como el impacto de haber sido entrevistado en sucesivas ocasiones anteriores (Garrido y Herrero, 2006; Manzanero, 2001; Masip y Garrido, 2007).

Respecto a la lista de validez, sucede lo mismo que con la entrevista: el mayor problema es que muchas veces no se utiliza. Además, tampoco hay una regla sobre el peso que cada elemento de la lista de validez debe tener sobre la decisión final (Masip y Garrido, 2007). Vrij (2005) señala diversos aspectos de la lista de validez que pueden ser problemáticos.

- a) El criterio *sugestionabilidad*. Como Vrij (2005) indica, es complicado medir la sugestionabilidad de un niño, y por lo tanto determinar si su declaración ha sido de alguna forma influenciada por un adulto.
- b) El criterio *emociones durante la entrevista*. Según este criterio, si el menor ofrece detalles del episodio de abuso sin mostrar ningún signo de

emoción o mostrando una emoción inapropiada, la declaración puede ser falsa. Pero el problema es que no se sabe con certeza lo que es un afecto apropiado.

- c) El criterio *declaraciones inconsistentes*. En este caso se sugiere que una declaración puede ser falsa si los entrevistados se contradicen a sí mismos en dos afirmaciones diferentes. Vrij plantea que es difícil determinar si una afirmación es inconsistente con otra, y que la investigación ha mostrado que la inconsistencia entre distintas afirmaciones no es un indicio válido de engaño.
- d) El criterio *falta de realismo*. Este criterio es difícil de valorar, ya que pueden darse casos en los que ciertos elementos incluidos en la historia hagan que ésta aparezca irreal, pero no obstante puede tratarse de una alegación verdadera (véase Vrij, 2005).
- e) Dificultad de determinar con precisión el impacto de un factor externo. Aún en aquellos casos en los que los expertos saben que alguno de los factores externos que aparecen en la lista de validez pueden estar presentes, es difícil determinar el impacto exacto de ese factor sobre las puntuaciones del CBCA.
- f) Utilización inapropiada de la lista de validez. Vrij indica que la lista de validez podría utilizarse de forma inadecuada si los expertos no consideran de qué modo un factor externo está afectando específicamente al menor que está siendo entrevistado. Es decir, no se trata solamente de establecer qué factores externos hay que tener siempre en cuenta, sino de determinar la forma diferencial en la que cada uno de ellos afecta a un determinado menor.
- g) Desestimación de la lista de validez. Finalmente, según señala el autor, a menudo los expertos toman las decisiones solamente en base a los resultados del CBCA, sin tener en cuenta la lista de validez.

El CBCA

El empleo del CBCA como instrumento para diagnosticar la veracidad de una declaración tiene varias limitaciones. El primer problema es de índole metodológica, y se refiere a que, actualmente, no hay un conjunto de reglas “oficiales” sobre cómo debe puntuarse exactamente la presencia o ausencia de cada criterio en la declaración. Como hemos indicado anteriormente, se han propuesto distintos métodos de puntuación, pero ninguno se ha constituido como una regla oficial. Así pues, distintos estudios experimentales han utilizado distintas formas de cuantificar la presencia o ausencia de criterios en la declaración. Por ejemplo, Akehurst, Bull, Vrij y Köhnken (2004) emplearon una escala tipo Likert de 5 puntos, mientras que otros (Horowitz et al., 1997; Lamb et al., 1997; Zaparniuk, Yuille y Taylor, 1995) emplearon un criterio dicotómico (presencia vs. ausencia), y Sporer (1997) empleó una escala de tres puntos (ausente, alguna indicación de presente y claramente presente).

Otro problema se refiere a que tampoco existe una “puntuación de corte” a partir de la cual se pueda afirmar con rotundidad que la declaración se corresponde con la realidad. Los investigadores que han analizado el CBCA coinciden en señalar que no hay una norma o regla de decisión oficialmente establecida sobre cuántos criterios y cuáles han de estar presentes para considerar que la declaración es verdadera (Manzanero, 2001; Raskin y Esplin, 1991a; Ruby y Brigham, 1997; Steller y Boychuk, 1992; Steller y Köhnken, 1990/1994; Vrij, 2000; Vrij y Akehurst, 1998). Tal como indican Masip y Alonso (2006), los métodos para separar las declaraciones verdaderas de las falsas que se han utilizado han sido diversos: (a) procedimientos estadísticos, tales como el análisis discriminante o de regresión; (b) aplicación de reglas de decisión, que hacen referencia a qué combinaciones específicas de criterios deben estar presentes para calificar una declaración como verdadera; y (c) tomar una decisión basándose en la opinión de evaluadores expertos en el CBCA.

Zaparniuk et al. (1995) realizaron un estudio en el que compararon la eficacia de dos reglas de decisión comúnmente utilizadas en distintos estudios para calificar una declaración como verdadera: a) presencia de los cinco

primeros criterios y al menos otros dos cuales quiera de los restantes, y b) Presencia de los tres primeros criterios y al menos otros cuatro criterios cualesquiera de los restantes. En los resultados de este estudio se encontró una tendencia que apuntaba a la superioridad de la primera de estas reglas, con la que se clasificaron correctamente el 78% de las declaraciones, sobre la segunda, con la que se alcanzó un 73% de precisión. No obstante, según los autores esta tendencia no es lo suficientemente significativa como para decantarse por un método definitivo de valoración.

Al margen de estas cuestiones de método, el principal problema del CBCA se refiere a su limitada precisión a la hora de diagnosticar una declaración como verdadera o falsa. En distintos estudios se ha examinado tanto la capacidad discriminativa de cada criterio a nivel individual como del CBCA en su conjunto, o la fiabilidad interjueces a la hora de identificar los criterios (por ejemplo, Akehurst et al., 2004; Horowitz et al., 1997; Köhnken et al., 1995; Lamb et al., 1997; Vrij, Edward, Roberts y Bull, 2000; Vrij, Akehurst, Soukara y Bull, 2004; Zaparniuk et al., 1995). El problema a la hora de comparar los resultados de estos estudios es que en muchos casos son muy diferentes entre sí, tanto respecto al número de criterios que se han analizado, como al tipo de participantes (se ha trabajado con individuos de muy diferentes edades) o el tema de la declaración. A pesar de estos inconvenientes, Vrij (2005) realizó una revisión de 37 estudios empíricos. Al analizar el porcentaje de estudios en los que cada criterio aparecía significativamente más en las declaraciones verdaderas que en las falsas, el autor descubrió que había profundas diferencias en la validez de los distintos criterios, tal como aparece en la Tabla 18.

Ante estos datos, Vrij concluye que hay una serie de criterios (elaboración no-estructurada, cantidad de detalles, engranaje contextual y reproducción de conversaciones) que presentan unos mejores niveles de discriminación que el resto. Por otro lado, hay criterios que muestran muy bajo poder discriminativo, como incomprensión de detalles, plantear dudas y sobre todo autodesaprobación, que fue examinado en seis estudios y en ninguno discriminó correctamente, y sí en sentido opuesto en dos de ellos (Vrij, 2005).

Tabla 18. *Porcentaje de Estudios en que Cada Criterio Aparecía más en Declaraciones Verdaderas y que Mostraron Buena Fiabilidad Interjueces (Vrij, 2005).*

Criterio	Porcentaje de estudios	
	Mayor presencia del indicio en verdades	Buena fiabilidad interjueces
1. Estructura lógica	53	79
2. Elaboración no-estructurada	64	50
3. Cantidad de detalles	80	77
4. Engranaje contextual	69	77
5. Descripción de interacciones	53	75
6. Reproducción de conversaciones	69	82
7. Complicaciones inesperadas	33	60
8. Detalles inusuales	53	62
9. Detalles superfluos	35	64
10. Incomprensión de detalles	12	86
11. Asociaciones externas relacionadas	40	63
12. Alusiones al propio estado mental	40	70
13. Atribución del estado mental del autor del delito	36	73
14. Correcciones espontáneas	35	46
15. Admisión de falta de memoria	46	58
16. Plantear dudas sobre el propio testimonio	18	80
17. Autodesaprobación	0	100
18. Perdonar al autor del delito	40	100
19. Detalles característicos del delito	50	75

Respecto a la precisión global del CBCA, Vrij (2005) encontró que ésta oscila en un rango bastante amplio, entre el 65% y el 90%, lo cual puede ser un reflejo de las diferencias en las condiciones de los estudios analizados. La precisión media para las verdades fue del 73% y para las mentiras del 72%. Un interesante hallazgo de este trabajo es que no se encontraron diferencias en precisión cuando las declaraciones las hacían niños o adultos, lo cual habla a favor de la aplicación del CBCA a ambos tipos de poblaciones.

Finalmente, respecto a la fiabilidad interjueces, para todos los criterios salvo elaboración no-estructurada, correcciones espontáneas y admisión de falta de memoria ésta fue buena en al menos un 60% de los estudios. Vrij consideró que la fiabilidad era buena si ésta superaba el valor .60, con independencia del índice empleado. Un resultado interesante es que en tres estudios en los que se analizó la fiabilidad para el CBCA en su conjunto en vez de para cada criterio por separado, las puntuaciones fueron mayores (78, 78 y 85, respectivamente), lo cual apunta a que la fiabilidad del CBCA total es mejor que la de sus indicios por separado.

En definitiva, el CBCA muestra una buena fiabilidad como instrumento, y la precisión en la discriminación de verdades y mentiras está significativamente por encima del azar. Sin embargo, unas tasas del 73% para las verdades y del 72% para las mentiras nos parecen demasiado bajas como para tomar decisiones definitivas sobre la veracidad de un declarante en juicios reales.

No obstante, tal y como ha planteado Vrij (2005), una buena forma de establecer una valoración del CBCA como instrumento científico susceptible de emplearse en casos reales es comprobar si cumple los “criterios *Daubert*”, promulgados en Estados Unidos para determinar los requisitos que debe cumplir la evidencia científica para ser admitida en los juicios. Para ello, se plantean cinco interrogantes: (a) ¿es científicamente comprobable la hipótesis en la que se basa la evidencia?; (b) ¿ha sido comprobada la proposición?; (c) ¿se conoce la tasa o probabilidad de error?; (d) ¿ha sido sometida la hipótesis y/o técnica a revisión por pares y publicación?; y (e) la teoría en la que se basa la hipótesis y/o técnica ¿está aceptada de forma general por la correspondiente comunidad científica? Según Vrij (2005), la hipótesis es científicamente comprobable (a) y ya ha sido comprobada (aunque mayormente en estudios de laboratorio) (b); ha sido sometida a revisión por pares y hay un número creciente de publicaciones (d). El hecho de que no esté suficientemente clara la tasa de error (ya que solamente se ha comprobado en estudios de laboratorio en donde se ha encontrado una tasa en torno al 30%, pero no se conoce en casos reales) (c) y no haya una postura unánime en la comunidad científica respecto a la teoría sobre la que se sustenta la hipótesis (e), determinan que no se cumplan suficientemente los criterios *Daubert* como para que la SVA sea admitida como evidencia científica en juicios reales.

Una vez descrita la SVA, que es la aproximación verbal que se ha empleado más y que más investigación ha generado (por lo que se le ha dedicado más espacio en este trabajo), en el siguiente apartado se presentan de forma breve otras herramientas que también pueden emplearse en la detección de la mentira desde la aproximación verbal.

Otras herramientas para la detección del engaño desde la aproximación del contenido verbal del discurso

La aproximación del Control de la Realidad (RM)

La *Reality Monitoring* o aproximación del Control de la Realidad parte de los planteamientos de Johnson y Raye (1981). Estas autoras proponen que tanto la percepción de estímulos externos como los propios pensamientos generan recuerdos, y al proceso por el cual una persona atribuye un recuerdo a una fuente externa (experiencia realmente vivida) o interna (experiencia imaginada o pensada, pero no realmente vivida) lo denominan *reality monitoring*.

Johnson y Raye (1981) señalan que los recuerdos pueden contener cuatro clases diferentes de información: (a) aspectos sensoriales como colores, formas o sonidos; (b) aspectos contextuales, referidos al tiempo o momento y al espacio o lugar; (c) aspectos semánticos; y (d) operaciones cognitivas, es decir, procesos cognitivos implicados en la elaboración de la información.

Lo que Johnson y Raye (1981) plantean es que una discriminación en la cantidad de cada uno de los tipos de información anteriores puede ayudar a distinguir entre recuerdos externamente generados y recuerdos imaginados (internamente generados). Así, se espera que las descripciones de recuerdos internamente generados contengan más operaciones cognitivas, ya que al ser la imaginación o el pensamiento el que crea el evento, habrán sido las operaciones cognitivas las implicadas en la creación y elaboración del recuerdo. Por otro lado, se espera que los recuerdos externamente generados, al haber sido vivenciados, contengan más información de tipo perceptual, contextual y semántica, por ejemplo referida al lugar en donde ocurrió el hecho, el momento, colores, olores, ruidos y otros detalles perceptuales (Johnson y Raye, 1981).

Teniendo en cuenta todo lo anterior, el RM podría aplicarse a la detección de la mentira, ya que, tal como han planteado algunos investigadores (véanse, por ejemplo, Alonso-Quecuty, 1994; Masip, Sporer, et al., 2005) hacer una declaración falsa (mentir) no es sino describir un evento internamente generado (con lo que se esperarí que tuviera más operaciones cognitivas), y hacer una

declaración verdadera es describir un recuerdo externamente generado (con lo que se esperaría una mayor cantidad de información contextual, sensorial y semántica).

Así pues, se han establecido una serie de criterios para determinar el tipo de información que contiene cada declaración. El conjunto de criterios utilizados ha variado tanto a lo largo de su desarrollo como dependiendo del investigador específico que ha empleado esta técnica (Sporer, 2004; Vrij, 2000). El lector interesado en el origen y desarrollo de la lista de criterios a la que aquí nos referiremos (el JMCQ: *Judgment of Memory Characteristics Questionnaire*, en español Cuestionario de Juicios de las Características del Recuerdo) puede consultar distintas fuentes (por ejemplo Masip y Alonso, 2006; véase también Sporer, 2004). El JMCQ está compuesto por ocho escalas o criterios que incluyen un total de 43 ítems (ver Anexo B).

La efectividad de cada criterio a nivel individual, así como la capacidad de la RM en su conjunto para discriminar recuerdos de hechos reales (verdades) de recuerdos de hechos imaginados (mentiras) ha sido comprobada en distintos estudios. Las conclusiones más importantes se resumen a continuación.

En lo que se refiere a la capacidad de discriminación de los criterios, Sporer (2004) señala dos importantes dificultades a la hora de comparar los distintos estudios. Primera, distintos autores han utilizado distintos grupos de criterios. Segunda, incluso en aquellos estudios en los que se han examinado los mismos criterios, éstos han sido operacionalizados de forma diferente. No obstante, se ha observado que, coincidiendo con los planteamientos de Johnson y Raye (1981), la información espacial, temporal y el realismo están más presentes en los recuerdos de hechos realmente vividos que en los recuerdos de hechos imaginados (véanse las revisiones Masip y Garrido, 2003; Masip, Sporer, et al., 2005; Sporer, 2004; Vrij, 2000).

Los resultados referidos a otros criterios como la información sensorial no son unánimes. Mientras que Vrij (2000), examinando distintos estudios, indica que la información sensorial discrimina de forma adecuada, en otras revisiones (Masip, Sporer, et al., 2005 y en español Masip y Garrido, 2003) se concluye

que este criterio no siempre permite discriminar correctamente. Por su parte, Sporer (2004) encuentra que cuando los individuos analizaban sus propias declaraciones, este tipo de información aparecía más en las que eran verdad que en las falsas, pero cuando los participantes analizaban las declaraciones de otros, el criterio discriminaba en sentido opuesto, es decir, aparecía más en las declaraciones falsas. En un estudio posterior Sporer y Sharman (2006) encontraron idénticos resultados respecto a la información sensorial. Quizá, tal como Masip y Alonso (2006) plantean, los resultados respecto a la información sensorial son contradictorios porque los distintos estudios han definido este criterio de forma diferente.

Finalmente, la investigación respecto a las operaciones cognitivas, hipotéticamente asociadas a las declaraciones falsas, muestra que este criterio no discrimina adecuadamente entre verdades y mentiras (Masip, Sporer, et al., 2005; Vrij, 2000). Quizá, tal como Sporer (2004) indica, haya que diferenciar entre distintos tipos de operaciones y procesos cognitivos, ya que, según este autor, no todos tienen porqué ser un indicio de que la declaración es falsa. En este sentido, Vrij et al. (2000) plantean que algunas operaciones cognitivas (por ejemplo, inferencias hechas a partir de distintos detalles que se recuerdan) aparecen como facilitadores del recuerdo de algunos aspectos en narraciones de hechos que realmente se han vivido.

Respecto a la precisión global del RM, la mayoría de los estudios han empleado la técnica del análisis discriminante o la regresión logística para obtener las tasas de precisión. Sporer (1997) obtuvo una tasa global de precisión del 71% (75% para las verdades y 68% para las mentiras). Vrij et al. (2000) hallaron unas tasas de precisión del 71% para las verdades y el 64% para las mentiras. Posteriormente Vrij et al. (2004) obtuvieron una precisión global del 77% (81% para las verdades y 73% para las mentiras) con una muestra de estudiantes universitarios de edades en torno a 22 años. En esta misma investigación, cuando los participantes eran más jóvenes (entre 14 y 15 años), la precisión mostrada fue un poco mayor, 82% a nivel global (92% para las verdades y 69% para las mentiras). En otra investigación, Sporer y Sharman (2006) encontraron una precisión media del 60% y un fuerte sesgo de veracidad (82% para verdades y 40% para mentiras).

Finalmente, los resultados de dos recientes revisiones sobre la RM (que reflejan de forma más global la precisión de esta herramienta) van en la misma línea que los estudios anteriores; Sporer (2004) concluye que la tasa media se establece en un rango entre el 64% y el 86% (entre el 61% y el 91% para las verdades y entre el 61% y el 85% para las mentiras) y Masip, Sporer, et al. (2005), tras revisar varias investigaciones, indican una precisión media del 72%, tanto para las declaraciones verdaderas como para las falsas.

¿CBCA o RM?

Vrij (2000) indicaba tres limitaciones importantes del CBCA que apuntaban a la necesidad de desarrollar una nueva herramienta que los superase. En primer lugar, aunque las tasas de precisión están por encima del nivel del azar, el método está aún lejos de alcanzar unos niveles aceptables para emplearlo en casos de juicios reales. En segundo lugar, el CBCA es un método de verificación de la verdad, no un instrumento para evaluar la veracidad. Por último, el CBCA estaría en principio contextualmente limitado, ya que inicialmente fue desarrollado para trabajar con casos de abusos sexuales. Así pues, algunos investigadores han visto en el RM un método alternativo al CBCA.

No obstante, lo cierto es que el RM no ha superado completamente estas tres limitaciones del CBCA. Quizá la más importante, la precisión en la discriminación de verdades y mentiras, sigue siendo la asignatura pendiente. Los estudios que han comparado el poder discriminativo de las dos técnicas han hallado unas tasas similares. Sporer (1997) obtuvo una precisión global para el RM del 71% (75% para las verdades y 68% para las mentiras) y una precisión global para el CBCA del 65% (70% para verdades y 60% para mentiras). Vrij et al. (2004) encontraron también una leve superioridad del RM sobre el CBCA (69% y 61% respectivamente); sin embargo, en un estudio anterior, Vrij et al. (2000) habían hallado una superioridad del CBCA sobre el RM (73% y 67% respectivamente). Estos resultados no aclaran qué técnica es la más discriminativa, y cuestionan el que el RM haya superado al CBCA en precisión.

Respecto a las otras dos limitaciones, sí es cierto que el RM ha sido creado como un instrumento de discriminación de declaraciones, y no de confirmación. En este sentido, el RM incluye un criterio (operaciones cognitivas) que se espera que aparezca más en declaraciones falsas. Sin embargo, la mayoría de criterios son criterios de verdad, y la investigación muestra que las operaciones cognitivas no permiten discriminar entre verdades y mentiras. Por último, el RM superaría la limitación contextual del CBCA, en el sentido de que no ha sido elaborado para analizar un tipo o tema específico de recuerdo. Otra ventaja del RM sobre el CBCA es, tal como Masip y Alonso (2006) indican, que cuenta con una base teórica más sólida y mejor articulada. Por otra parte, el entrenamiento necesario para manejar el RM es mucho más simple y breve.

En conclusión, si a través de la investigación y el trabajo con el RM se consigue en un futuro depurar sus criterios de forma que se mejoren los niveles de precisión, esta técnica, por su base teórica más sólida y su mayor facilidad de entrenamiento y aplicación, puede superar al CBCA.

Las Aberdeen Report Judgment Scales (ARJS)

En el apartado anterior se han observado distintos estudios que comparaban el CBCA con el RM. Un resultado interesante que ofrecen todos ellos es que la precisión conjunta de las dos aproximaciones es superior a la precisión para cualquiera de ellas por separado (Sporer, 1997; Vrij et al., 2004; Vrij et al., 2000) Este hallazgo sirvió de base a Sporer (1997) para plantear la posibilidad de integrar el CBCA y la RM en un solo instrumento para la detección del engaño.

La herramienta resultante de esta integración se beneficiaría de un adecuado soporte teórico (perspectiva cognitiva propia del RM) desde el que se han derivado distintos criterios, así como de la presencia de los criterios del CBCA, más socialmente contextualizados, que se centran en las interacciones entre los personajes que protagonizan el episodio que se está analizando. Tal como Sporer (1997) indica, el RM aportaría criterios de naturaleza más cognitiva, mientras que, en el ámbito forense en el que se ha aplicado el CBCA, se subraya la importancia de la interacción entre víctima y perpetrador, lo cual

se ve reflejado en criterios del CBCA que son de naturaleza más social, como interacciones verbales y no verbales o reproducción de conversaciones. Como el propio autor indica, estaríamos ante un instrumento integrado de naturaleza cognitivo-social.

A partir de estos planteamientos, Sporer (1997) en primer lugar estableció correlaciones entre los criterios del CBCA y el RM. En este análisis encontró un importante número de correlaciones significativas, tal como se presenta en la Tabla 19, lo cual muestra que, efectivamente, hay un buen grado de solapamiento entre los dos conjuntos de criterios.

Seguidamente, el autor realizó un análisis factorial con los dos grupos de criterios introducidos conjuntamente. Los resultados de este análisis mostraron dos posibles soluciones, una de cinco factores y otra de seis. Sporer señala la solución de cinco factores como más deseable dado que era más fácilmente interpretable. Tales factores eran: (a) Consistencia lógica y realismo, (b) cantidad de detalles y engranaje contextual, (c) sentimientos y pensamientos, (d) claridad y (e) interacciones verbales y no verbales. Estos cinco factores explicaron un total del 61.4% de la varianza, y a partir de ellos Sporer desarrollaría las *Aberdeen Report Judgment Scales* (ARJS).

Las ARJS están compuestas por un total de 52 ítems distribuidos en 13 escalas (por ejemplo, Masip y Alonso, 2006; Sporer, 2004; Sporer et al., 2000), tal como se indica en el Anexo C. La presencia en una declaración de cada tipo de información a la que se refiere cada ítem se puntúa en una escala que va de 1 (*poco presente*) a 7 (*muy presente*).

Tabla 19. *Correlaciones Significativas Entre Criterios del CBCA y el RM (Sporer, 1997).*

Criterios CBCA	Criterios RM							
	Claridad viveza	Información sensorial	Información espacial	Información temporal	Reconstruibilidad	Sentimientos	Operaciones cognitivas	Realismo
Consistencia lógica	0.40**				0.69**			0.72**
Producción no estructurada	-0.39**				-0.42**			
Cantidad de detalles		0.26*	0.33**	0.30**			0.31*	
Engranaje contextual			0.29**	0.24*	0.27	0.25*	0.26*	0.33**
Interacciones			0.33*				0.28*	
Conversaciones			0.32**					
Complicaciones					-0.28*			
Detalles inusuales					-0.39**			-0.75**
Detalles superfluos			0.27*					
Procesos psicológicos propios	0.28*					0.68**	0.63**	
Procesos psicológicos del agresor						0.35**		

*p < .05; **p < .01.

¿Son las ARJS un buen instrumento para diferenciar verdades de mentiras?

El propio Sporer ha realizado distintos estudios en los que ha comprobado la utilidad de las AJRS como instrumento de detección del engaño. Sporer et al. (2000), en uno de los estudios iniciales con las ARJS, trataron de comprobar tanto la consistencia interna del instrumento como la fiabilidad inter-evaluadores. Para ello entrenaron a tres personas para que puntuasen 160 declaraciones (la mitad verdaderas y la otra mitad falsas) con los 52 ítems de las ARJS. Cada uno de los ítems se puntuó en una escala de 1 (*poco presente*) a 7 (*muy presente*). En la Tabla 20 se presentan los resultados respecto a la fiabilidad y la consistencia interna.

Tabla 20. *Media de la Correlación Inter-jueces (r) y Consistencia Interna de las ARJS (Sporer et al., 2000).*

Escalas	<i>r</i>	Alfa de Cronbach
1. Realismo y coherencia	0.28	0.84
2. Claridad y viveza	0.55	0.70
3. Detalles centrales y periféricos y precisión de los detalles	0.69	0.91
4. Información espacial	0.54	0.84
5. Información temporal	0.82	0.69
6. Impresiones sensoriales	0.71	0.18
7. Emociones y sentimientos	0.60	0.90
8. Pensamientos y procesos cognitivos	0.61	0.73
9. Procesos de memoria y ensayo	0.54	0.63
10. Interacciones verbales y no verbales	0.58	0.74
11. Detalles extraordinarios	0.33	0.45
12. Falta de de deseabilidad social	0.43	0.48
13. Significación personal e implicaciones	0.39	0.71
TOTAL ARJS	0.71	0.91

Podemos ver que, aunque algunas escalas (detalles centrales y periféricos y precisión de los detalles, información temporal e impresiones sensoriales) mostraron una buena fiabilidad, otras (realismo y coherencia, detalles extraordinarios, falta de de deseabilidad social y significación personal e implicaciones) mostraron unas puntuaciones demasiado bajas. En cualquier caso, la fiabilidad global del instrumento (0.71) es buena. En relación con la consistencia interna también se observan diferencias entre distintas escalas. Aunque la mayoría mostraron un buen índice, en algunos casos (impresiones sensoriales, detalles extraordinarios y falta de de deseabilidad social) la puntuación fue demasiado baja. En cualquier caso, al igual que la fiabilidad, la

consistencia interna global del instrumento (0.91) fue muy buena. Ante estos resultados, Sporer et al. (2000) concluyen que las ARJS muestran unas propiedades psicométricas aceptables para su utilización. No obstante, es necesario depurar esta herramienta, sobre todo a nivel de cada ítem individual, en donde las diferencias en las puntuaciones mostradas entre las escalas se agravan mucho. Una posible explicación es que quizá el entrenamiento no fue suficiente, y serían necesarias más horas de trabajo. En cualquier caso, los autores están de acuerdo en que algunos ítems necesitan ser redefinidos para mejorar las puntuaciones referidas al acuerdo inter-jueces.

Respecto a la precisión de las ARJS, Sporer (2004) presenta un estudio en el que se analizaron las transcripciones de declaraciones verdaderas y falsas de un grupo de soldados sobre su participación en unas maniobras militares. Según indica el autor, a pesar del breve entrenamiento que recibieron los jueces encargados de puntuar las declaraciones, la fiabilidad alcanzada fue en general satisfactoria, excepto para las escalas de realismo y coherencia y detalles extraordinarios. Por otra parte, las escalas de información temporal, emociones y sentimientos, detalles extraordinarios y falta de deseabilidad social diferenciaron significativamente las declaraciones verdaderas de falsas. Un análisis discriminante produjo un porcentaje de clasificaciones correctas del 69% (77.1% para las mentiras y 62.9% para las verdades).

En un estudio más reciente, Barnier, Sharman, McKay y Sporer (2005), utilizando un análisis discriminante obtuvieron una precisión global para las ARJS del 64.1% (58.9% para las verdades y 66.7% para las mentiras). Adicionalmente, se pidió a la persona que puntuó cada criterio de las ARJS en cada declaración que emitiera un juicio de veracidad sobre cada historia. La precisión global de estos juicios fue del 62.2% (52.2% para las verdades y 67.8% para las mentiras)¹².

¹² Este estudio es más complejo, y se manipularon distintas variables (que no describiremos aquí por razones de espacio) como la valencia de las declaraciones (si describían episodios positivos o negativos) y el tipo de mentiras (eventos imaginados [se pedía al declarante que hiciera la declaración como si hubiera imaginado el evento] o mentiras [se pedía al declarante que hiciese la declaración con la intención de engañar al entrevistador]). No obstante, el lector interesado puede consultar el artículo de Barnier et al. (2004).

Teniendo en cuenta los datos presentados, las ARJS se presentan como un instrumento prometedor para la detección de engaño. Si bien es cierto que la herramienta necesita aún ser perfeccionada, y en ello sigue trabajando Sporer y su equipo, es posible que las tasas de precisión lleguen a superar las alcanzadas con el CBCA y el RM. Además, las ARJS pueden sin duda llegar a ser un instrumento más deseable que los anteriores, dado su carácter integrador cognitivo-social.

El Linguistic Inquiry and Word Count (LIWC)

El *Linguistic Inquiry and Word Count* (LIWC), diseñado en el 2001 por Pennebaker, Francis y Booth, no es en sí mismo un instrumento o protocolo específicamente construido para la detección de la verdad o la mentira. Se trata de un software diseñado para analizar textos palabra por palabra y clasificarlas por categorías¹³. En concreto, el LIWC toma cada palabra del texto y la compara con las del diccionario interno del programa, las cuales están organizadas en categorías lingüísticas. De esta forma, a medida que el programa va encontrando palabras del texto que coinciden con palabras de una determinada categoría del diccionario, las va contabilizando. En la salida de resultados, el LIWC ofrece información sobre la frecuencia de 72 categorías lingüísticas diferentes, con lo que permite observar qué tipo de palabras aparecen más frecuentemente un texto (Pennebaker, Francis y Booth, 2001). Este programa se ha utilizado en diversos campos de la psicología, sobre todo en temas de salud, tratando de analizar el discurso de enfermos de distintas dolencias (por ejemplo, Suriá, 2004).

Dentro del campo de la detección del engaño, algunas investigaciones han utilizado el LIWC para analizar las diferencias lingüísticas en declaraciones verdaderas y falsas. Newman, Pennebaker, Berry y Richards (2003) emplearon el programa en cinco estudios diferentes. En los estudios 1, 2 y 3 el tema de las declaraciones verdaderas o falsas fue la opinión del declarante hacia el aborto.

¹³ El lector interesado en una descripción más detallada de cómo se ha elaborado este programa informático y las listas de palabras que componen las categorías del diccionario que incluye, puede acudir a la propia web en la que se comercializa (véase "La aplicación del LIWC", sin fechar, <http://www.liwc.net/liwcspanol/liwcdescription.php>), en donde se ofrece información del producto y desde la que se puede adquirir.

En el Estudio 4, los declarantes tuvieron que hacer una descripción verdadera y otra falsa de personas que realmente les gustaban o no. En el Estudio 5, se estableció un paradigma de falso robo, en el que tanto quienes habían cometido el robo como quienes no tuvieron que negar haberlo cometido. Se determinó que los temas de las declaraciones fueran diferentes con el objetivo de establecer un perfil general del tipo de palabras cuya frecuencia de aparición fuera distinta al mentir y decir la verdad independientemente del tema de la declaración.

Se plantearon tres hipótesis. En primer lugar, se esperaba que las declaraciones falsas tuvieran menos autorreferencias, más palabras que denotaran emociones negativas y menos indicadores de complejidad cognitiva y verbos que expresaran movimiento (por ejemplo: andar, ir, etc.). Una segunda hipótesis predecía que el perfil lingüístico hallado a partir de las declaraciones de un tema específico sería generalizable a declaraciones sobre temas diferentes. Finalmente, la tercera hipótesis predecía que el perfil general de las declaraciones falsas creado a partir de los cinco estudios alcanzaría mayor precisión en la detección del engaño que los juicios de personas no entrenadas.

Una vez que las declaraciones se transcribieron y se analizaron con el LIWC, se trató de discriminar entre verdades y mentiras en cada estudio utilizando un perfil de declaración falsa creado a partir de los datos de los estudios restantes¹⁴. Por ejemplo, las declaraciones del Estudio 1 se clasificaron como verdades o mentiras a partir de un perfil de declaración falsa creado con los datos de los estudios 2, 3, 4 y 5, y del mismo modo se procedió con los restantes estudios. Los resultados referidos a la precisión obtenida se muestran en la Tabla 21.

¹⁴ Un perfil se definía así como un modelo general integrado por los elementos lingüísticos que era más probable que constituyeran una mentira. Éste se construía a partir de las características de las declaraciones falsas de todos los estudios excepto las de aquel que se iba a examinar (véase Newman et al., 2003, para una explicación más detallada).

Tabla 21. *Precisión en Cada Estudio de la Investigación de Newman et al. (2003).*

	Precisión para las mentiras	Precisión para las verdades	Precisión global
Estudio 1	60%	58%	59%*
Estudio 2	61%	57%	59%*
Estudio 3	66%	69%	67%**
Estudio 4	52%	54%	53%
Estudio 5	53%	43%	48%

* $p < .05$. ** $p < .001$ respecto a una comparación con el azar.

Respecto a la utilidad de las distintas categorías para detectar mentiras, el mejor predictor del engaño fue la categoría palabras excluyentes (menos presentes en declaraciones falsas), que discriminó en los cinco estudios, seguida de los verbos de movimiento (más en las declaraciones falsas), que discriminó en los estudios 1, 2, 3 y 5. Los autores consideraron también buenos predictores a las categorías emociones negativas (aparecieron más al mentir), que discriminó adecuadamente en los estudios 1, 2 y 4, pronombres en primera persona (menos empleados en las declaraciones falsas), que discriminó en los estudios 3, 4 y 5, y pronombres en tercera persona (más presentes en declaraciones falsas), que discriminó adecuadamente en los estudios 3 y 5.

Estas cinco categorías se consideraron como perfil general de una declaración falsa, y se introdujeron como predictores de mentiras en un análisis de regresión logística en el que se consideraron conjuntamente las declaraciones de los cinco estudios. Los resultados arrojaron una precisión del 59% para las mentiras y del 62% para las verdades. El porcentaje global de clasificaciones correctas fue del 61%, significativamente superior al azar.

Finalmente, se comparó la precisión que este perfil general alcanzó para las declaraciones de los tres primeros estudios tomados conjuntamente (68% para las mentiras, 66% para las verdades y 67% a nivel global) con la precisión de los juicios de un conjunto de individuos que no habían recibido ningún tipo de entrenamiento (30% para las mentiras, 74% para las verdades y 52% a nivel global). Estos resultados son quizá los más interesantes de esta investigación, en la medida en que se pueden concluir dos ventajas del LIWC sobre los juicios humanos: en primer lugar, la precisión media obtenida con este programa es superior a la precisión de personas no entrenadas, y en segundo lugar, parece eliminarse el fuerte sesgo hacia la verdad de los juicios humanos.

En un estudio posterior, Hancock, Curry, Goorha y Woodworth (2004) diseñaron un paradigma experimental en el que dos personas conversaban por escrito de forma interactiva a través de un programa informático. A una de estas personas se le pidió que mintiera o dijera la verdad durante el intercambio de mensajes. A la otra persona se le indicó simplemente que siguiese la conversación. Estos autores utilizaron el LIWC para analizar los cambios en el lenguaje de ambos participantes (del emisor al mentir o decir la verdad y del receptor al ser o no mentido). Se utilizaron las siguientes categorías: cantidad de palabras, pronombres personales de primera, segunda o tercera persona, negaciones, palabras referidas a emociones, palabras referidas a sensaciones y palabras excluyentes. Adicionalmente, se examinó la cantidad de preguntas, tanto del emisor como del receptor.

Los resultados mostraron que, en general, cuando los participantes mentían producían más palabras, utilizaban más referencias a otros (pronombres de tercera persona) y empleaban más palabras referidas a sensaciones (por ejemplo ver, oír, sentir, etc.) que cuando decían la verdad. Respecto al estilo lingüístico de los receptores, cuando éstos eran mentidos, utilizaban más palabras, aunque sus frases eran más cortas, presentaban más palabras referidas a sensaciones y hacían más preguntas a su pareja conversacional que cuando les decían la verdad.

Como se puede observar, a excepción de las referencias a otras personas, las demás categorías que mostraron cambios significativos al mentir en este estudio no coinciden con las que se mostraron como buenos predictores del engaño en el estudio de Newman et al. (2003). No obstante, no sería adecuado comparar los resultados de estos dos estudios dadas las grandes diferencias en el contexto experimental de ambos. Una limitación de esta investigación es que en los paradigmas de estudio del engaño en un contexto de interacción hay muchas variables que no se pueden controlar y que son propias de la idiosincrasia de cada conversación. Por ejemplo, las respuestas a preguntas, matizaciones o aclaraciones que el individuo que escucha la mentira o verdad hace al emisor son diferentes en cada pareja conversacional. Esto, sin duda, complica mucho la posibilidad de comparar los datos de cada interacción con los de las demás.

Bond y Lee (2005a) realizaron una interesante investigación con el LIWC. Observaron que los criterios que se han empleado en investigaciones en las que se ha trabajado con el RM son similares a algunas de las categorías del LIWC, y se plantearon la posibilidad de contrastar el poder discriminativo de dichos criterios utilizando el LIWC en lugar de jueces humanos. Además, también trataron de comprobar la capacidad discriminativa del perfil general de las declaraciones falsas que Newman et al. (2003) habían encontrado (al que los investigadores se refieren como modelo NP)¹⁵ y compararon los modelos RM y NP con los juicios de veracidad efectuados por un grupo de presos.

Siguiendo las predicciones de la RM, se esperaba que en las declaraciones verdaderas aparecieran más palabras relacionadas con información espacial, temporal y sensorial, y más términos afectivos, pero menos palabras referidas a mecanismos cognitivos. Desde el modelo NP se esperaba que en las declaraciones verdaderas aparecieran más autorreferencias (pronombres personales de primera persona), más referencias a otros (pronombres personales de tercera persona), más palabras excluyentes, menos palabras referidas a emociones negativas y menos verbos de movimiento. Se transcribieron 76 declaraciones verdaderas y 76 falsas y se codificaron con el LIWC.

Tras someter a un análisis de regresión logística las categorías del modelo RM, se halló un efecto significativo para las palabras sensoriales (tal como predecía el modelo) y también un efecto significativo para las palabras espaciales (pero en dirección opuesta a lo que predecía el modelo, por lo que esta categoría se excluyó a la hora de examinar la precisión del LIWC con las categorías del RM). No se encontraron efectos significativos para la información temporal ni para palabras referidas a procesos cognitivos.

Utilizando las categorías del RM información sensorial y perceptual, información temporal, información afectiva y mecanismos cognitivos, el LIWC clasificó correctamente el 71.1% de las mentiras y el 64.5% de las verdades. Hubo dos categorías del modelo NP que discriminaron significativamente

¹⁵ La denominación NP se deriva de tomar las iniciales de los dos primeros autores (Newman y Pennebaker) del estudio de Newman et al. (2003).

verdades de mentiras: verbos de movimiento y referencias a otras personas. Considerando únicamente estas dos categorías, se alcanzó un nivel de aciertos del 69.7% y una tasa de rechazos correctos del 68.4%. Análisis adicionales mostraron que la precisión alcanzada desde el modelo RM no se diferenció significativamente de la alcanzada con el modelo NP, y ambas fueron superiores al nivel del azar.

Estos resultados se compararon con la precisión que obtuvieron un grupo de presos jóvenes (de edades entre 18 y 27 años) y otro de presos mayores (65 a 84 años) que juzgaron las declaraciones basándose en su intuición. El grupo de presos jóvenes obtuvo una precisión para las mentiras del 71.1%, que no se diferenció de la precisión obtenida desde los modelos RM y NP. La precisión para las verdades, sin embargo, fue del 36.8%, significativamente inferior a la obtenida desde los modelos RM y NP. El grupo de presos mayores obtuvo una precisión para las mentiras igual al nivel de azar (50.0%) e inferior a los modelos RM y NP, y una precisión para las verdades mucho mayor (81.6%). Observando estos resultados, se aprecia (al igual que en el estudio de Newman et al., 2003) que la principal ventaja del empleo del LIWC es que se evitan los sesgos (bien hacia la mentira o hacia la verdad) que aparecen cuando los jueces son humanos. Por otra parte, a diferencia de dicho estudio, en este caso la precisión en la detección del engaño fue similar con el empleo del LIWC que sin él. No obstante, hay que tener en cuenta que en otros estudios se ha observado que los presos han obtenido mejores tasas de precisión que los no-presos (Aamodt y Custer, 2006; Bond, Malloy, Arias, Nunn y Thompson, 2005), o que las creencias de los presos sobre los indicios del engaño son más precisas que las de los no presos (Granhag et al., 2004).

En definitiva, es muy interesante la informatización de las distintas técnicas de detección del engaño, dado que se observa que se ven libres de los sesgos característicos de los juicios humanos. Además, podría evitarse pasar por todo el proceso de entrenamiento de los jueces a la hora de codificar la presencia o ausencia de distintos criterios, así como posibles problemas derivados de la fiabilidad inter-jueces. No obstante, no se puede obviar que los resultados respecto a la precisión son aún susceptibles de mejora. Además, tal como han indicado Bond y Lee (2005a), el diccionario del LIWC debe ampliarse para, por

ejemplo, recoger con más precisión categorías de palabras que se correspondan con los criterios de distintos modelos, como el RM. No obstante, trabajos como el de Bond y Lee (2005a), que han utilizado el LIWC para trabajar desde los supuestos de modelos ya elaborados como la RM, son aún escasos, y es sin duda necesaria más investigación al respecto.

La técnica SCAN

La técnica SCAN (*Scientific Content Analysis*; en español “Análisis Científico del Contenido”) no se basa en una teoría científicamente sostenible, ni ha sido validada con estudios empíricos libres de importantes limitaciones metodológicas. Todo esto hace que nos cuestionemos seriamente su validez.

Esta herramienta se deriva de la intuición y experiencia como interrogador y poligrafista de Avinoam Sapir, miembro de la policía israelí y fundador de la compañía *Laboratory for Scientific Interrogation* (LSI), que actualmente se dedica al entrenamiento y la distribución comercial de esta técnica¹⁶. La razón por la que se incluye la técnica SCAN en esta Tesis es porque, a pesar de lo dicho, es una herramienta que utiliza la policía de países como Canadá, Australia, Bélgica, distintos estados de América o Israel para evaluar declaraciones en casos reales (Smith, 2001). Es posible que su uso se haya extendido debido a los impactantes eslóganes publicitarios que el LSI utiliza para captar clientes. En este sentido, Masip et al. (2002b) hacen un buen análisis crítico de la publicidad excesivamente pretenciosa e irrealista del LSI. Otros autores como Shearer (1999) presentan también críticas, más duras si cabe, en la misma línea.

La forma de utilizar la técnica SCAN para diferenciar verdades de mentiras consiste en aplicar una serie de criterios a una declaración que el propio sospechoso debe haber escrito de su puño y letra. En este sentido, se hace especial hincapié en que debe ser solamente el sospechoso el que, armado de papel y bolígrafo, escriba su declaración sobre los hechos, sin la intervención del interrogador, que en ningún caso ha de contaminar la narración con

¹⁶ El lector interesado puede dirigirse a la siguiente dirección web: <http://www.lsiscan.com/>

preguntas o indicaciones de ningún tipo (Adams, 1996; Driscoll, 1994; Lesce, 1990).

Nótese que estamos utilizando la palabra “sospechoso”, porque la técnica SCAN, al contrario que, por ejemplo, el CBCA, es una herramienta enfocada a verificar la falsedad de la declaración. De esta forma, la presencia de un criterio en la declaración apunta a la falsedad de la misma. De hecho, algunos profesionales que se dedican a entrenar policías en el uso de esta técnica, como la agente Susan H. Adams, definen la técnica SCAN específicamente como un medio para obtener una confesión (Adams, 1996).

Adams (1996) indica que la aplicación de esta herramienta supone dos pasos: en primer lugar, se establece la “norma” o suposición de cómo una declaración verdadera debe ser. En segundo lugar se aplican los criterios para determinar si la declaración se desvía de la norma; si es así, se concluye que es falsa. En el Anexo D se presenta una lista de algunos de los criterios de la técnica SCAN (no hay una lista exhaustiva de tales criterios), así como una breve explicación de cada uno¹⁷.

Apenas existen estudios que hayan intentado validar estos criterios, y adolecen de serias limitaciones metodológicas. Driscoll (1994) aplicó diez criterios sobre las transcripciones de treinta declaraciones verdaderas o falsas. A pesar de que el autor, al final del estudio, concluye que la técnica es capaz “aunque con límites” de diferenciar verdades de mentiras (se clasificó correctamente el 72% de las verdades y el 94% de las mentiras), las debilidades metodológicas que tiene este estudio son tales, que sus resultados simplemente no pueden tenerse en cuenta.

Algunas de estas limitaciones se refieren a la forma en que se estableció la verdad de base de las declaraciones, que solamente permitió al autor hablar de declaraciones “aparentemente ciertas” o “dudosas” y no de verdades y mentiras propiamente dichas. Además, la cantidad de declaraciones

¹⁷ Por razones de espacio, el Anexo D solamente presenta una breve descripción de cada criterio. No obstante, el lector interesado puede consultar diversas fuentes (Adams 1996; Driscoll, 1994; Masip et al., 2002b; Smith, 2001). En todos estos trabajos se presenta una descripción más detallada, con ejemplos de cada criterio.

aparentemente ciertas ($n = 11$) y dudosas ($n = 19$) examinadas fue desigual. Así mismo, sólo una persona analizó las declaraciones, y esta persona sabía cuáles eran aparentemente ciertas y cuales dudosas. La limitación más grave es que el autor no utilizó ninguna prueba estadística para examinar la significatividad de las diferencias, sino que, tal como expresan Masip et al. (2002b), hizo una “inspección visual” de los datos. Estos y otros problemas (véase Masip et al., 2002b, para una crítica más detallada de este estudio) hacen, como hemos dicho, que los resultados de Driscoll (1994) no sean fiables.

Smith (2001) realizó un estudio en el que se planteó diversos objetivos: (a) valorar la efectividad de la técnica SCAN para identificar declaraciones verdaderas y falsas (es decir, declaraciones sobre las que no se tenía seguridad de que fueran verdaderas o falsas), (b) comprobar si policías entrenados con la técnica SCAN obtenían mejor precisión en sus juicios que policías no entrenados que utilizaran su experiencia e intuición personal, (c) evaluar si distintos niveles de entrenamiento con la técnica SCAN se correspondían con distintos niveles de precisión en los juicios y (d) comprobar si algunos de los criterios de la técnica SCAN eran mejores discriminadores que otros.

Se contó con un total de 25 declaraciones de casos reales referidas a distintos delitos. Tras aplicar las pruebas para establecer la verdad de base¹⁸, se concluyó que 20 eran mentiras, 4 eran verdades y 3 inconclusivas. Cinco grupos de observadores juzgaron la veracidad de este grupo de declaraciones. El Grupo 1 incluía tres policías, quienes en su trabajo policial solo de vez en cuando utilizaban la técnica SCAN. El Grupo 2 estaba compuesto por dos policías que muy raramente utilizaban esta técnica. El Grupo 3 estaba formado por policías que utilizaban la técnica SCAN frecuentemente. El Grupo 4 lo integraban dos detectives veteranos en el cuerpo que no utilizaban la técnica SCAN y el Grupo 5 estaba compuesto por tres policías noveles, recientemente

¹⁸ Para que una declaración se considerara verdadera, ésta tenía que cumplir al menos dos de los siguientes criterios: evidencia inequívoca a favor de la veracidad del testimonio, detención de otra persona, confesión de otra persona, condena de otra persona. Para que una declaración se considerase falsa, tenía que cumplir al menos dos de los siguientes criterios: confesión del declarante, arresto del mismo, que la policía hubiese abandonado el caso y evidencia inequívoca a favor de la falsedad de la declaración.

ingresados en el cuerpo. Cada grupo actuó como un solo participante, es decir, emitió un solo juicio por declaración, excepto el Grupo 4, en el que cada uno de los dos policías que lo formaban emitió un juicio individual.

Todos los grupos obtuvieron unos resultados por encima de lo esperado por azar. La precisión de los cinco grupos para las verdades osciló entre el 80% y el 100%. La precisión para las mentiras osciló entre el 65% y el 100% para los grupos 1, 2, 3 y 4. El Grupo 5 (policías recién ingresados en el cuerpo) obtuvo una precisión del 45%, sensiblemente inferior.

Entre los grupos 1, 2 y 3, que utilizaban la técnica SCAN de forma diferencial (de vez en cuando, muy raramente, frecuentemente) no hubo diferencias significativas en precisión (85.2%, 74.1% y 81.5%, respectivamente) es decir, no se encontró un efecto de la práctica. La precisión conjunta de estos tres grupos (80.2%), fue mejor que la de los policías recién ingresados en el cuerpo (48.2%), pero no hubo diferencias significativas con respecto a los policías veteranos (59.2% para el policía 1 y 66.7% para el policía 2) que no utilizaban la técnica SCAN y habían basado sus juicios en su intuición y experiencia (véase Smith, 2001).

Los criterios más utilizados por los grupos que habían recibido entrenamiento con la técnica SCAN fueron: uso inadecuado de pronombres, falta de memoria y cambios en el lenguaje. Entre los menos utilizados estuvieron las correcciones espontaneas y las emociones. En cualquier caso, el autor indica que hubo poca consistencia entre los grupos en el uso de los criterios. Respecto a su utilidad real para detectar mentiras, solamente los criterios cambios en el tiempo verbal, no negar la acusación y el uso de conexiones o enlaces innecesarios se encontraron significativamente más en las declaraciones falsas que en las verdaderas.

Smith, aunque califica su propio estudio como “preliminar”, concluye que la técnica SCAN es una herramienta que muestra un buen potencial para la discriminación de declaraciones verdaderas y falsas, y considera que entrenar a los policías en el uso de esta herramienta puede ayudarles a detectar el engaño.

Aunque esta investigación se realizó siguiendo unas pautas metodológicas más adecuadas que el estudio de Driscoll (1994) sigue habiendo algunos aspectos que hacen que los resultados no sean concluyentes. En primer lugar, de las 25 declaraciones presentadas, 20 eran mentiras, 4 eran verdades y 3 inconclusivas. Ante esta distribución de la veracidad real de las declaraciones, las tasas de precisión que se encontraron no tienen porqué indicar la precisión de los policías. Podrían perfectamente reflejar una tendencia de estos a juzgar las declaraciones como falsas, resultado que otras investigaciones han encontrado (véase Masip et al., 2002b).

En segundo lugar, Los grupos 1, 2 y 3 (compuestos por policías que utilizaban la técnica SCAN de forma diferencial) eran tan pequeños que no pueden ser representativos de las distintas poblaciones que pretendían representar.

En tercer lugar, dentro de cada uno de estos grupos, los juicios se emitieron por consenso entre los integrantes. Hubiera sido más deseable que cada uno de los miembros del grupo emitiese sus juicios individualmente y luego hallar unas puntuaciones promedio de cada grupo, sobre todo teniendo en cuenta que estos tres grupos se comparan luego con el Grupo 4, en el que los juicios sí se emitieron de forma individual y no por consenso.

En cuarto lugar, el hecho de que solamente se encontrasen diferencias significativas en precisión entre los grupos entrenados (1, 2 y 3) tomados conjuntamente y el grupo 5 (policías recién ingresados en el cuerpo), puede reflejar que dichas diferencias se deben a la experiencia adquirida en los años de servicio como policía, y no al entrenamiento recibido en la técnica SCAN. Esta posibilidad se vería apoyada por el hecho de que entre los grupos 1, 2 y 3, en los que todos eran policías veteranos, no hubo diferencias en las tasas de precisión a pesar del distinto grado de práctica. Además, entre estos tres grupos y el grupo 4, formado por policías también veteranos, tampoco hubo diferencias, a pesar de que estos hicieron sus juicios sobre la base de su experiencia e intuición y no de la técnica SCAN. Estas y otras limitaciones (véase Masip et al., 2002b, para un análisis más detallado) hacen que nos cuestionemos los resultados de este estudio.

Finalmente, Porter y Yuille (1996) realizaron un experimento que sí reúne los suficientes requisitos metodológicos. Se utilizaron declaraciones basadas en un delito simulado que los propios autores habían preparado. El problema es que solamente se analizaron tres criterios de la técnica SCAN: Conectores innecesarios, uso inadecuado de pronombres (en concreto se observó la frecuencia de las desviaciones de la primera persona), y ratio de los componentes de la declaración (definido como distribución temporal de los hechos). Se halló que ninguno de estos tres criterios discriminó significativamente entre declaraciones verdaderas y falsas.

En definitiva, podríamos extraer las siguientes conclusiones sobre la técnica SCAN: (a) Sus criterios no se apoyan en ninguna teoría o paradigma científico claramente estructurado, sino que más bien se derivan de la intuición y experiencia personal de su creador; (b) hay muy poca investigación que haya tratado de validar esta herramienta; (c) de los tres estudios que hemos encontrado al respecto, solamente los resultados de Driscoll (1994) ofrecen un apoyo empírico claro a la técnica SCAN, los resultados de Smith (2001) ofrecen un apoyo limitado y los de Porter y Yuille (1996) ningún apoyo; y (d) curiosamente, parece haber una relación entre la calidad metodológica de estos estudios y los resultados hallados: a mayor calidad metodológica, menor apoyo empírico.

Todas estas razones hacen que no se pueda concluir que la técnica SCAN sea un instrumento adecuado para la detección del engaño. Por lo tanto, cabe cuestionarse las afirmaciones triunfalistas sobre la calidad y precisión de esta herramienta emitidas por la empresa que la distribuye. Todos los profesionales que por su trabajo podrían beneficiarse de un buen protocolo para la evaluación de la veracidad deberían consultar la literatura científica existente sobre la fiabilidad y validez de cualquier técnica antes de utilizarla. Todo ello redundaría, sin duda, en un trabajo responsable y de una buena calidad.

Aplicación de la informática a la detección verbal del engaño

No queremos finalizar este apartado sin hacer referencia a una serie de trabajos que, desde la aproximación verbal de la detección del engaño, se han

centrado, por un lado, en el desarrollo de programas informáticos que permitan una “detección automática”, y por otro lado, en la detección de mentiras transmitidas a través del ordenador. Por ejemplo, Zhou, Burgoon, Nunamaker y Twitchell (2004) trataron de construir un *software* informático para identificar características morfológicas, sintácticas y léxico-semánticas del lenguaje que permitieran distinguir verdades de mentiras en correos electrónicos. Se hipotetizó que los mensajes de los emisores que mintieran, en comparación con los de quienes dijeran la verdad y con los de los receptores, mostrarían mayor cantidad de elementos lingüísticos (palabras, verbos, frases...), denotarían más expresividad (entendida como la ratio de adjetivos y adverbios vs. nombres y verbos), más emociones positivas y estarían redactados de forma más informal (mayor empleo de abreviaturas tipográficas). Además, se hipotetizó que los mensajes falsos mostrarían falta de seguridad, no inmediaticidad (caracterizada por el empleo del pasivo, referencias grupales, términos generalizadores, etc.), menor complejidad, menor diversidad léxica y menor especificidad del lenguaje (información espacio-temporal y perceptual).

Se analizaron los *e-mails* de 14 díadas en que el emisor había sido sincero y de 16 en que había mentido. Los resultados ofrecieron un apoyo limitado a las hipótesis. Los mensajes de los emisores que mentían puntuaban significativamente más en cantidad, falta de seguridad, emociones positivas e informalidad (en este último caso con una significación marginal), y menos en diversidad que los mensajes de los emisores veraces. Los *e-mails* falsos también puntuaban significativamente más en falta de seguridad, expresividad y afecto negativo, y marginalmente menos en diversidad que los mensajes de los receptores.

En un trabajo posterior, Qin, Burgoon y Nunamaker (2004) analizaron las declaraciones de 48 personas, 25 de las cuales habían cometido un falso robo, siendo las otras 23 inocentes. De los culpables, 16 fueron entrevistados vía *chat* y nueve por audio-conferencia. De los inocentes, 12 fueron entrevistados por *chat* y 11 por audio-conferencia. Se analizó lingüísticamente la transcripción de las grabaciones de audio junto con los textos del *chat*. Se examinaron 19 características verbales relacionadas con la calidad, la complejidad, la especificidad, la expresividad y la informalidad del lenguaje. Las

características lingüísticas significativas fueron longitud de las frases ($p = .008$), promedio de palabras por frase ($p = .021$) e informalidad, definida por el número de errores en el texto marcados por el *Word Perfect* ($p = .037$). Los autores también consideraron relevantes otras características que se acercaron a la significación estadística, como el número de palabras ($p = .096$), la FKGL¹⁹ ($p = .056$) y la complejidad de la frase ($p = .082$). A continuación se diseñaron distintos árboles de decisión para clasificar informáticamente las declaraciones en función de sus características lingüísticas. Cuando se introdujeron las 19 características iniciales tanto de las entrevistas vía *chat* como de las realizadas por vídeo-conferencia, se obtuvo una precisión del 58.30% al clasificar los textos como verdaderos o falsos. Cuando se introdujeron solamente las seis claves significativas o marginalmente significativas la precisión aumentó hasta el 62.50%. Cuando se introdujeron estas claves sólo en la condición *chat*, la precisión se situó en el 78.58%. Estos autores han seguido trabajando en esta línea de investigación, y Bugoon (2004, en Masip y Alonso, 2006) afirma haber alcanzado una precisión de entre el 61% y el 91% al juzgar las declaraciones verdaderas y entre el 61% y el 88% al juzgar las declaraciones falsas con el *software* que están desarrollando.

Otros trabajos están profundizando más en las posibilidades de investigación en este campo. Por ejemplo Zhou y Zhang (2006) han estudiado cómo puede influir el tamaño del grupo de comunicadores (díadas o tríadas) en los indicios de engaño que aparecen en las comunicaciones a través del ordenador. Encontraron que, efectivamente, aparecían indicios diferentes en los mensajes falsos en función de si eran dos o tres las personas que se comunicaban simultáneamente. En concreto, en interacciones diádicas las comunicaciones falsas presentaban una menor complejidad del lenguaje que las comunicaciones verdaderas ($p = .001$). De forma marginalmente significativa, los mentirosos mostraban una menor participación ($p = .060$) y complacencia²⁰ ($p = .071$) en sus mensajes que los veraces. En interacciones triádicas, quienes mentían iniciaban más temas de conversación ($p = .047$) y mostraban una

¹⁹ Se trata de un componente de la complejidad al que los autores se refieren como *Flesch-Kincaid Grave Level*.

²⁰ La palabra que define el indicio en inglés es *pleasantness*. Los autores operacionalizan el indicio como el grado de simpatía y lo agradable que resulta el mensaje (véase Zhou y Zhang, 2006).

menor participación ($p = .016$) que sus compañeros sinceros, y sus mensajes reflejaban una mayor complejidad cognitiva ($p = .005$).

Los trabajos que se acaban de presentar son una muestra de cómo la investigación está tratando de avanzar en la informatización de la detección del engaño a partir de indicadores verbales. Sería interesante poder desarrollar un *software* informático que permita analizar este tipo de mensajes, y los resultados a los que Bugoon (2004, en Masip y Alonso, 2006) hace referencia son prometedores. Además, en los estudios expuestos se aborda un nuevo campo de estudio del engaño. Se trata de detectar mentiras en una forma de interacción cada vez más extendida, la comunicación a través del ordenador (vía *chat* o *e-mail*). No obstante, aún es necesaria más investigación.

Conclusiones

En este apartado hemos presentado algunos métodos para la detección del engaño desde la aproximación verbal. La primera y más importante conclusión que se podría extraer al respecto es que, al igual ocurre con la aproximación no verbal, tampoco existe desde esta perspectiva ningún indicio o protocolo que sea totalmente fiable.

Además, el meta-análisis de DePaulo, Lindsay, et al. (2003) indica que aunque existen conductas verbales probabilísticamente asociadas al engaño, éstas no son muchas. De los indicios que pudieron examinarse en más de cinco estudios, solamente siete discriminaron significativamente verdades de mentiras con un tamaño del efecto mayor que $.20$ ó menor que $-.20$ ²¹ (detalles [-.30], estructura lógica [-.25], plausibilidad [-.23], frases negativas y quejas [.21], impresiones de inmediatez verbal y vocal [-.55], impresiones de falta de certeza verbal y vocal [.30] e implicación verbal y vocal [-.21]). Otras cuatro conductas se mostraron como buenas herramientas de discriminación (admisiones de falta de memoria [-.42], asociaciones externas relacionadas [.35], inmediatez verbal [-.31] y corregir espontáneamente la declaración [-.29]). No obstante, estos últimos indicios solamente se comprobaron en cinco estudios o menos. Se necesita pues, investigación adicional. Finalmente, hubo

²¹ Recordemos que en un tamaño del efecto positivo asociaba el indicio al engaño, y un tamaño del efecto negativo a la verdad.

un pequeño grupo de claves con un tamaño del efecto mayor que .20 en valores absolutos, pero que se había podido calcular con precisión sólo en uno o dos estudios (modificadores [-.52], admisión de incertidumbre [-.63], estilo de discurso relacionado con el tema [-.87], razones argumentadas para no recordar [-.75], alusiones a la responsabilidad de la conducta [.34], y repetición de elementos de la declaración [-.65]). La falta de evidencia experimental no permite considerarlos como buenas herramientas de detección.

En cualquier caso, lo cierto es que la detección del engaño a partir de las palabras ofrece protocolos más estandarizados (el CBCA, el RM, la técnica SCAN) que la aproximación no verbal. En consecuencia, no sería adecuado comparar los niveles de precisión obtenidos desde las aproximaciones no verbal y verbal, ya que en el primer caso se trata de la precisión de individuos que basan sus juicios en su mera intuición, mientras que en el segundo siempre han recibido algún tipo de información o entrenamiento. Por ejemplo, en cualquier estudio del CBCA siempre se da algún tipo de instrucciones o formación a los evaluadores sobre los criterios y cómo utilizarlos.

Es importante en este punto hacer una llamada a la prudencia a la hora de utilizar cualquiera de los protocolos descritos en casos reales, sobre todo si hablamos de técnicas como la SCAN, que no tienen una base científica. Aún en el caso de métodos que sí se derivan de la investigación científica (SVA y RM), las tasas de error deben, a nuestro juicio, reducirse más para que su empleo sea aceptado en los juzgados. Quizá la informatización de tales procedimientos sea un buen camino para reducir algunos sesgos y errores propios de su utilización.

APROXIMACIÓN BASADA EN INDICADORES PSICOFISIOLÓGICOS

Se ha dejado para el final esta orientación porque es la que más se aleja del contenido y perspectiva general de esta Tesis. No obstante, dado que el polígrafo (instrumento “estrella” desde esta vía de investigación) ha generado y sigue generando una gran controversia respecto su utilidad, y dado que se utiliza en investigaciones criminales en países como Bélgica, Canadá, Israel, Japón, Turquía, Singapur, Corea del Sur, Méjico, Paquistán, Filipinas, Taiwán,

Tailandia y algunos estados de EEUU (Bull et al., 2004), es necesario examinar de cerca tanto su funcionamiento como la base teórico-científica en la que se sustenta.

Desde la aproximación psicofisiológica de la detección del engaño se parte de la premisa de que cuando un individuo miente experimenta *arousal*, lo que va a provocar que manifieste un conjunto de respuestas psicofisiológicas que se pueden medir o registrar mediante instrumentos diseñados a tal efecto, como por ejemplo el polígrafo (Masip y Garrido, 1999; Vrij, 1998b, 2000). Así pues, estrictamente hablando, el polígrafo no es un detector de mentiras, sino un detector de ciertas respuestas psicofisiológicas que, desde esta perspectiva, se hipotetizan asociadas al acto de mentir.

Cómo funciona el polígrafo

Tal como se describe en distintos trabajos (por ejemplo, Vrij, 2000) el polígrafo es un instrumento que actúa en tres fases: medición, amplificación y representación gráfica. En concreto, a través de unos sensores colocados en diferentes partes del cuerpo se recogen distintos tipos de actividad fisiológica mientras el individuo es sometido a varios tipos de preguntas. Las señales recogidas por los sensores, pasan directamente a un aparato que funciona como un amplificador de las mismas, que a su vez está conectado bien a unas agujas que imprimen en forma de gráficos sobre una tira de papel los distintos tipos de actividad fisiológica recogida, o bien a un ordenador que muestra los registros gráficos en la pantalla. En definitiva, el polígrafo mide la actividad psicofisiológica humana durante un periodo de tiempo y ante distintos tipos de estímulos. Como indican Masip y Alonso (2006), la actividad fisiológica que se recoge es la respuesta psicogalvánica, la presión sanguínea relativa, el ritmo cardiaco, y la respiración.

Pruebas poligráficas para la detección de la mentira

La RIT (Relevant / Irrelevant Test)

Esta es una de las primeras y más sencillas pruebas que se han utilizado para saber cuándo alguien está mintiendo mediante el polígrafo. Una vez que

la persona que se quiere interrogar está conectada al polígrafo a través de los sensores, se le hacen dos tipos de preguntas: (a) Preguntas relevantes para el delito, que están relacionadas con el suceso sobre el que se quiere saber si el individuo miente; el examinador indica a la persona que va a ser interrogada que debe contestar “no” a este tipo de preguntas, ya que lo contrario sería admitir haber cometido el delito; y (b) preguntas irrelevantes para el delito, es decir, que no guardan relación con el tema que se está investigando; se trata de cuestiones ante las que el examinador sabe con seguridad que el interrogado dirá la verdad (por ejemplo, preguntarle su nombre o el día de la semana). A continuación, el examinador compara la magnitud de las respuestas psicofisiológicas ante los distintos tipos de preguntas. Si la reacción psicofisiológica ante una pregunta relevante es más intensa que la reacción ante una pregunta irrelevante, se entiende que el individuo está mintiendo a la pregunta relevante.

Esta prueba aparece duramente criticada en diversos trabajos, por ejemplo en la revisión sobre el polígrafo y su aplicación que Bull et al. (2004) elaboraron para la *British Psychological Society* (BPS). Tal como indican estos autores, el miedo o la ansiedad de una persona inocente a ser erróneamente declarada culpable podrían desencadenar una reacción más fuerte al enfrentarse a una pregunta relevante (al percibir la gravedad e importancia de la misma), que ante una pregunta irrelevante, la cual el individuo puede afrontar con más tranquilidad al observar que no está relacionada con el caso del que es sospechoso. Dado que, según Bull et al. (2004), hay un acuerdo general entre los distintos expertos en el polígrafo sobre no utilizar esta prueba, no nos extenderemos más en ella.

La CQT (Control Question Test²²)

Esta es la prueba que normalmente se utiliza cuando se juzgan delitos reales en aquellos estados de los EE.UU. donde los resultados del polígrafo son admitidos como prueba en los tribunales (Masip y Alonso, 2006). En la

²² Otros autores se refieren a esta prueba como *Comparison Question Test* (Honts, 2004). En esta Tesis consideramos equivalentes las dos denominaciones, dado que ambas se refieren a la misma prueba, y nos referiremos a ella con las siglas CQT.

aplicación de la CQT se manejan tres tipos de preguntas: irrelevantes²³, relevantes y control (véanse por ejemplo Honts, 2004; Lykken, 1998; Masip, 2002; Vrij, 1998b, 2000, para una descripción más detallada de esta prueba y su aplicación). Las preguntas irrelevantes son generales, no relacionadas con el caso que se está tratando y no se espera que generen ningún *arousal* ni, por lo tanto, respuesta psicofisiológica a considerar. Según Vrij (2000), éstas se utilizan como relleno, o para comprobar, por ejemplo, si el examinado está prestando atención. Serían preguntas sobre el nombre del individuo, su nacionalidad o su edad. Las preguntas relevantes se refieren específicamente al caso que se está tratando. Se espera que, ante ellas, la respuesta sea siempre negativa (lo contrario sería asumir la responsabilidad del delito o crimen que se esté investigando). Se hipotetiza que los culpables experimentarán más *arousal* que los inocentes, ya que los primeros son conscientes de haber cometido el delito, mientras que los segundos no. Algunas preguntas de este tipo serían, por ejemplo, las referidas directamente a la autoría del delito o la forma en que éste se cometió.

Las preguntas control se refieren a temas que no están directamente asociados con el caso que se estudia, sino con episodios o sucesos de la vida del interrogado que, sin revestir la gravedad y especificidad del delito que se trata, aún así guardan un cierto paralelismo. Este tipo de preguntas se presentan siempre en un formato muy general, deliberadamente vago y ambiguo, y generalmente se refieren a largos periodos de tiempo de la vida del interrogado. Por ejemplo, si el caso que se está tratando implicase un delito de robo, una pregunta control podría ser: “¿ha tomado alguna vez en su vida algo que no era suyo?” Si el caso implicara un asesinato, una pregunta control podría ser: “¿alguna vez ha hecho daño o herido a alguien a lo largo de su vida?”. Uno de los aspectos más delicados cuando se presenta este tipo de preguntas, es que el interrogador debe tener la habilidad de hacer que tanto los interrogados culpables como los inocentes contesten de forma negativa (aún sabiendo que casi todas las personas alguna vez hemos hecho daño a alguien). Si alguien alegase en esta dirección y pretendiese contestar de forma

²³ Algunos autores, como Vrij (2000), también se han referido a este tipo de preguntas como preguntas neutrales.

afirmativa, el interrogador debe ser capaz de convencerle para que no lo haga (por ejemplo, reformulando la pregunta de manera ligeramente distinta). En general, el argumento que se suele utilizar es indicar al interrogado que (siguiendo nuestro ejemplo) si se descubre que ha hecho daño a alguien alguna vez, un jurado podría pensar que es el tipo de persona capaz de haber cometido el delito del que es sospechoso. El interrogador trata de esta forma de que tanto los individuos inocentes como los culpables experimenten un cierto *arousal*.

Según los defensores de esta prueba, los inocentes mostrarán más *arousal* ante una pregunta control, dado que no pueden estar seguros de que dicen la verdad, que ante una relevante, a la cual contestarán la verdad dado que saben que son inocentes. Por el contrario, los culpables experimentarán más *arousal* ante una pregunta relevante que ante una control, dado que aunque mienten a las dos, son conscientes de que la pregunta relevante implica una mayor gravedad y peores consecuencias para ellos que la pregunta control.

Raskin (1990/1994) define distintas fases en la aplicación de la prueba CQT:

Entrevista pretest. Se trata de una entrevista previa a la prueba poligráfica que dura al menos una hora. Raskin (1990/1994) señala que el poligrafista, antes de reunirse con la persona que va a ser examinada, debe haberse familiarizado con los hechos del caso. Al inicio de la entrevista, se debe indicar al examinado que todo el proceso va a ser grabado, seguidamente hay que informarle sobre el propósito y temas de la prueba, de sus derechos legales, y hay que pedirle que firme un formulario de consentimiento. Los objetivos que se persiguen con la entrevista son los siguientes:

- a) Establecer *rapport* con el examinado. Según Raskin, una pequeña charla que permita al examinado exponer el incidente desde su punto de vista, además de clarificar posibles ambigüedades ayudará a establecer un clima de confianza.
- b) Obtener información de base. Se trata de descubrir cualquier problema físico o mental que pueda afectar a la validez de la prueba. Para ello se revisa el historial médico y psiquiátrico del examinado. También se

examinan sus actitudes hacia el engaño y el examen poligráfico y se le pregunta sobre cualquier dato biográfico que pueda ser relevante.

- c) Desarrollar y revisar las preguntas de la prueba. El poligrafista desarrolla las preguntas para la prueba posterior a partir de la información que ha recabado. Dichas preguntas se presentan y discuten con el examinado. El objetivo que se persigue es doble: asegurarse de que el interrogado entiende las preguntas y explicarle que las respuestas a las mismas deben ser un simple “no” sin más explicaciones.
- d) Convencer al interrogado de la infalibilidad del polígrafo. El poligrafista debe ser capaz de convencer al interrogado de que el polígrafo es muy preciso y que, sin duda, será capaz de detectar cualquier mentira que diga. Para ello se emplea lo que se conoce como *stimulation test* (prueba de estimulación). Para esta prueba se suele utilizar un test de números (véase Raskin, 1990/1994) o una baraja de cartas (véase Vrij, 2000). El examinado selecciona un número o una carta de la baraja (dependiendo del método empleado). El examinador adivina el número o la carta elegida a partir de la observación de la respuesta fisiológica mostrada ante la presentación de distintos números o cartas a los que el entrevistado ha respondido siempre “no”. El poligrafista insiste entonces en la efectividad del polígrafo²⁴.
- e) Generar miedo en el culpable. Según Raskin (1990/1994), al terminar la demostración anterior, el poligrafista debe informar al interrogado de que ya conoce las características de sus reacciones fisiológicas al mentir y al decir la verdad, de forma que no tendrá ningún problema con la prueba poligráfica si es completamente sincero, pero que cualquier mentira será clara e inequívocamente detectada. El objetivo es generar miedo en los culpables, lo cual disparará el *arousal* al enfrentarse a una pregunta relevante.

²⁴ Vrij (2000) indica que aunque en la mayoría de los casos el entrevistador tiene éxito en esta prueba, un error podría ser fatal, ya que pondría en entredicho la capacidad del polígrafo, y probablemente afectaría negativamente al resto de la prueba. Es por ello que, según este autor, los entrevistadores a veces “hacen trampas”, tales como marcar la carta elegida por el entrevistado.

Administración de la prueba poligráfica. El examinador va haciendo las preguntas y el examinado contesta “sí” (ante las preguntas irrelevantes) o “no” (ante las preguntas relevantes y control). Raskin indica que la secuencia de preguntas debe repetirse al menos tres veces, rotando el orden de las preguntas neutras y control para evitar que el interrogado tenga reacciones anticipatorias debidas a que espera que las preguntas se presenten siempre en el mismo orden. Si los resultados no son concluyentes tras tres repeticiones, pueden hacerse dos repeticiones más (Raskin, 1990/1994).

Interpretación de la salida de datos del polígrafo. En general hay dos formas de interpretar los datos obtenidos: aproximación global y puntuación numérica. En la aproximación global, el examinador se forma una impresión general comparando informalmente las respuestas psicofisiológicas obtenidas ante las preguntas control y las preguntas relevantes. Para tomar una decisión final, la conclusión obtenida de esta comparación se coteja con el comportamiento del individuo durante la prueba poligráfica, así como con otro tipo de evidencia o pruebas obtenidas por otros medios diferentes del polígrafo. La aproximación numérica trata de minimizar la influencia de cualquier evidencia que no haya sido obtenida con el polígrafo. Se establece una comparación para cada canal entre las respuestas psicofisiológicas a las distintas preguntas relevantes y control (por ejemplo, se compara la pregunta control 1 con la pregunta relevante 1; la pregunta control 2 con la pregunta relevante 2; y así sucesivamente). Se contemplan entonces varias posibilidades: (a) si no hay diferencias entre respuestas psicofisiológicas (por ejemplo si las respuestas a la pregunta control 1 y relevante 1 son iguales), se asigna a ese par una puntuación de 0; (b) si la respuesta ante la pregunta relevante es claramente mayor, se asigna una puntuación de -1, si es muy notable, de -2, y si es extremadamente notable, de -3; (c) si la respuesta ante la pregunta relevante es claramente inferior se asigna una puntuación de 1, si es muy notable, de 2, y si es extremadamente notable, de 3 (Vrij, 2000). Finalmente, se suman las puntuaciones obtenidas para todos los canales en todas las preguntas en las tres repeticiones de la secuencia de preguntas. Según Raskin (1990/1994), si la puntuación final obtenida es -6 o menor, se concluye que el individuo no ha pasado la prueba del polígrafo; es decir, ha mentado. Si la puntuación final es 6

o mayor, se concluye que el individuo ha pasado la prueba; es decir, ha dicho la verdad. Si la puntuación final está entre 5 y -5, la prueba se califica como inconcluyente.

Tasas de precisión obtenidas con la CQT

En la Tabla 22 se presentan los resultados de distintas revisiones sobre la precisión de la CQT, tanto en estudios de laboratorio como en estudios de campo.

Tabla 22. *Resultados de Distintas Revisiones Sobre la Precisión de la CQT Expresados en Porcentajes Medios.*

Revisiones	Mentirosos			Veraces		
	Correct.	Inconclus.	Incorrect.	Correct.	Inconclus.	Incorrect.
<i>Laboratorio</i>						
Honts (2004)	82%		7%	83%	7%	10%
Vrij (2000)	73%	≈ 20%	9%	66%	≈ 20%	13%
<i>Campo</i>						
Bull et al. (2004) ^a	83 - 89%		10-17%	53-78%		20-47%
Honts (2004)	89%	10%	1%	59%	29%	12%
Vrij (2000)	87%	≈7%	10%	72%	7%	21%

^aLas tasas se ofrecen en intervalos, dado que el trabajo de Bull et al. (2004) revisa los resultados de distintas revisiones, no de estudios individuales.

Además de las revisiones presentadas en la Tabla 22, uno de los trabajos más importantes que hay que considerar a la hora de valorar la precisión del polígrafo es el informe del *National Research Council (2003)*²⁵. Se trata de un extenso informe elaborado con el fin de arrojar luz sobre el fundamento científico y la utilidad práctica del polígrafo. Aborda el análisis de la precisión de la CQT desde la Teoría de la Detección de Señales, por lo que la precisión se expresa como la mediana de A (área bajo la curva ROC)²⁶. La capacidad del

²⁵ Tal como se indica a comienzo del informe, el *National Research Council* se estableció en los Estados Unidos a partir de la Academia Nacional de las Ciencias en 1916 con el objetivo de aunar los esfuerzos de la comunidad científico tecnológica y el conocimiento académico para asesorar al gobierno federal. El NRC se ha convertido en el principal organismo, tanto de la Academia Nacional de las Ciencias como de la Academia Nacional de Ingeniería, a la hora de asesorar tanto al gobierno, como al Departamento de Energía y a la comunidad científica. El informe sobre el polígrafo al que nos referimos en este trabajo fue solicitado al NRC por el Departamento de Energía, y en su elaboración se implicaron 18 profesionales de distintos campos científicos.

²⁶ El rango de A abarca de 0 a 1. El valor de 0.5 se corresponde con el azar. Cuanto más se acerque A a 1, mejor será la discriminación.

polígrafo para distinguir verdades de mentiras en estudios de laboratorio fue $A = 0.85$, y en los estudios de campo $A = 0.89$.

Una primera conclusión que se puede extraer de estas revisiones sobre la CQT es que tanto en los estudios de campo como en los de laboratorio, se muestra una mejor precisión en la identificación de personas que mienten que al identificar individuos que dicen la verdad. En concreto, es especialmente preocupante la poca precisión a la hora de clasificar personas veraces en los estudios de campo. El problema de los falsos positivos (individuos veraces erróneamente clasificados como mendaces) es especialmente relevante si tenemos en cuenta que muchos de los datos de estos estudios proceden de casos reales en los que el resultado de la prueba puede haber sido vinculante a la hora de establecer un veredicto de culpabilidad o inocencia. Los resultados que acabamos de observar, junto con otros aspectos criticables de la prueba CQT, sugieren que esta prueba puede mostrar un sesgo hacia la mentira, tal como ha indicado Lykken (1998).

De las revisiones a las que nos hemos referido, consideramos especialmente relevantes las de Bull et al. (2004) y del NRC (2003), ya que en ambos casos se trata de informes oficiales sobre el polígrafo. Por ello, de acuerdo con Masip y Alonso (2006), se les presupone mayor neutralidad que a otros trabajos cuyo enfoque puede verse sesgado por el posicionamiento de sus autores hacia una u otra prueba poligráfica. En el caso del NRC (2003), como se ha dicho se trata de un extenso informe publicado en los Estados Unidos. El trabajo de Bull et al. (2004) es un informe realizado por un equipo de seis profesionales que la *British Psychological Society* (BPS) publicó en el Reino Unido en el año 2004. En comparación con el anterior, se trata en esta ocasión de un informe mucho más breve. Tanto el trabajo del NRC como el de Bull et al. (2004) ofrecen unas tasas de precisión que distan de la infalibilidad que los poligrafistas pretenden hacer creer a los individuos a los que les aplican la CQT.

Limitaciones y críticas a la CQT

La CQT ha sido criticada por diferentes motivos, especialmente por quienes abogan por otro tipo de pruebas poligráficas. En primer lugar, diversos autores han puesto en duda la asunción principal de la CQT de que, a diferencia de los culpables, los inocentes mostrarán mayores respuestas psicofisiológicas a las preguntas control que a las preguntas relevantes. Ekman (1985/1992) por ejemplo, propone cinco razones por las que algunas personas inocentes podrían presentar un patrón opuesto al hipotetizado (o, por lo menos, sus resultados serían inconcluyentes):

a) Es posible que algunas personas inocentes piensen que la policía ha cometido un error al sospechar de ellas y someterlas a la prueba poligráfica. A partir de esta idea, dichas personas podrían pensar que, al igual que la policía ha cometido un error con ellos, también el polígrafo, como método policial, podría equivocarse.

b) Una persona inocente, por el hecho de que la policía la considere sospechosa, o por cualquier otro tipo de experiencia o razón, puede considerar que la policía como Cuerpo, y los policías como funcionarios, son injustos en la realización de su trabajo. Es posible que esta persona pueda temer que el poligrafista también vaya a ser injusto con él.

c) El sospechoso inocente puede temer al polígrafo en sí mismo, ya que puede desconfiar de la infalibilidad absoluta de cualquier máquina, y por extensión de aquella de la que depende que sea juzgado culpable o inocente de un delito.

d) El sospechoso puede ser una persona temerosa o asustadiza por naturaleza, o puede sentirse culpable por el simple hecho de pensar que es el tipo de persona de la que la policía desconfía.

e) Es posible que aunque el individuo examinado sea realmente inocente, por distintas causas manifieste una fuerte reacción emocional ante sucesos relacionados con el delito. Por ejemplo, si se trata de un asesinato y no tenía una buena relación con la víctima, puede sentir remordimientos, si fue el

primero en ver el cadáver y ello le causó una fuerte impresión, si el caso le evoca recuerdos angustiosos de otras situaciones, y un largo etc.

A estas cinco razones, Vrij (2000) añade un detalle adicional importante. Según este autor, hay numerosos libros y artículos en los que se expone la forma de trabajar del polígrafo y la aplicación de la CQT. Cualquiera de estas publicaciones describirá los aspectos esenciales y debilidades de esta prueba, las estrategias que el poligrafista utiliza para convencer al examinado de que la máquina es infalible y, sin duda, las tasas de precisión y de error. En palabras del propio autor: “la prueba poligráfica se volverá progresivamente menos efectiva en la medida en que los examinados no crean al examinador” (Vrij, 2000, p. 182). Cada persona es diferente y se enfrenta al polígrafo con unas circunstancias personales diferentes, las cuales, en algunos casos, influirán sin duda en el resultado de la prueba poligráfica. Tal como señala Lykken (1998), la premisa subyacente a la CQT no tiene en cuenta las diferencias individuales entre personas, con lo que es posible que funcione para algunos individuos, pero no para todo el mundo y en todos los casos.

Otra crítica que se ha hecho a la CQT se refiere al difícil rol y excesivo peso que tiene el examinador. El resultado de esta prueba depende en gran medida de las habilidades individuales que tenga el poligrafista a la hora de aplicarla. Como indican diversos autores (por ejemplo, Lykken, 1998; Vrij, 2000), en primer lugar el experimentador debe ser capaz de elegir preguntas control que generen una reacción fisiológica más fuerte que las preguntas relevantes en el examinado inocente. Por otro lado, en el caso de individuos culpables, las preguntas control seleccionadas deben generar menos *arousal* que las relevantes. Teniendo en cuenta que no hay un procedimiento estandarizado en la elección de las preguntas control, y que el examinador no debe saber a priori quién es culpable y quién inocente, elegir las preguntas control adecuadas no es una tarea fácil.

Si la elección de las preguntas control es compleja, más complicada aún es la tarea del examinador a la hora de convencer al examinado de que debe contestar negativamente a las mismas. Como se ha indicado, dado el carácter deliberadamente ambiguo de tales preguntas y el hecho de que abarquen

amplios periodos de tiempo, la mayor parte de los examinados dudan sobre su respuesta a este tipo de preguntas. Tal como indica Lykken (1998), el examinador debe hacer gala de todas sus habilidades para elaborar preguntas control ante las cuales el sospechoso acceda a contestar negativamente. La situación se complica más por el hecho de que, al mismo tiempo, el examinador debe ser capaz de convencer al examinado de la infalibilidad del polígrafo. Es decir, se crea una contradicción, ya que por un lado se crea una situación que empuja al examinado a mentir, y por el otro se le trata de convencer de que el polígrafo es infalible y detectará sin duda cualquier mentira que diga. Esta conducta es, cuanto menos, éticamente cuestionable, tal como Vrij (2000) señala. Finalmente, es posible que el examinador, por distintas circunstancias, tenga expectativas o creencias previas sobre la culpabilidad del sospechoso. De acuerdo con Vrij (2000), sin duda dichas expectativas o creencias previas influirán en los resultados del polígrafo.

Otra objeción a la CQT se refiere a la forma en que se puntúan los resultados. Respecto a la aproximación numérica, descrita anteriormente, Vrij (2000) pone de manifiesto la falta de reglas definidas para establecer cuando la diferencia entre dos respuestas (por ejemplo, relevante y control) es clara, notable, o extremadamente notable. Respecto a la aproximación global, Vrij (2000) señala que la forma en que se toma la decisión en este caso es muy subjetiva, depende demasiado del examinador, y no es verificable. Como ya se ha dicho, el examinador poseerá casi siempre información respecto al examinado y al caso, y tendrá sus propias creencias respecto a la culpabilidad o inocencia del sospechoso. Desde la aproximación global no hay ninguna forma de aislar la posible influencia de estos factores sobre la interpretación de los resultados. Finalmente, tampoco existe una regla o fórmula estandarizada que pueda asegurar que, ante unos mismos resultados, distintos evaluadores vayan a tomar las mismas decisiones.

Respecto al problema de los posibles sesgos a la hora de puntuar los resultados poligráficos, los defensores de la CQT proponen el empleo de procedimientos informáticos. Según Honts (2004), a lo largo de los últimos años se han desarrollado modelos y procedimientos estadísticos estandarizados fiables, así como formas computerizadas de interpretación de

las puntuaciones. Kircher y Raskin (2002) presentan con cierto detalle las posibilidades de aplicar programas informáticos a la interpretación de datos poligráficos, y se muestran optimistas respecto a las ventajas que tiene el ordenador tanto a la hora de recoger los datos poligráficos como a la hora de analizarlos. Ofrecen además evidencia de distintas investigaciones y concluyen que la aplicación del ordenador al polígrafo reduce posibles sesgos del examinador (véase Kircher y Raskin, 2002 para una explicación detallada de las ventajas de la aplicación de distintos protocolos informatizados a una prueba poligráfica). No obstante, los autores hacen hincapié en que el ordenador no debe en ningún caso sustituir totalmente al poligrafista humano ni tomar la decisión final, sino que debe ser más bien una herramienta de apoyo al servicio del profesional.

En los epígrafes anteriores se ha presentado la CQT y algunas de las críticas que ha suscitado entre diversos miembros de la comunidad científica. No obstante, esta prueba cuenta también con importantes defensores, hasta tal punto que, tal como se indica en Masip y Alonso (2006), la literatura sobre el polígrafo es en muchos casos un reflejo de la pugna entre defensores y opositores de la CQT. Descritas, pues, sus limitaciones más importantes, no sería objetivo finalizar este apartado sin expresar, en palabras de Honts (2004), la respuesta de alguien que defiende esta prueba ante las críticas anteriores:

El empleo de medidas psicofisiológicas para la detección del engaño en situaciones forenses está fuertemente apoyado por estudios de laboratorio y de campo de alta calidad. Aunque existe controversia sobre la aplicación forense del polígrafo, la mayor parte de los científicos informados en los Estados Unidos muestran opiniones positivas sobre la prueba de la pregunta de comparación (Honts, 2004, pp. 119-120).

Refiriéndose a los críticos de la CQT, este autor escribe:

Sus posiciones generalmente se caracterizan por la siguiente posición general: La racionalidad de la CQT es irracional, y por lo tanto la prueba no puede funcionar. Su trabajo generalmente se caracteriza por idear experimentos que demuestren que la CQT no funciona y que no es posible realizar investigación válida o ética, pero su postura carece generalmente de soporte empírico. Confían selectivamente en investigaciones realizadas con

una pobre calidad, pero no hacen referencia a estudios empíricos de alta calidad. En particular, desechan todos los estudios de laboratorio que tengan algún valor para establecer la validez de la CQT, aunque están deseosos de aceptar datos de laboratorio cuando estos muestran que la CQT falla (Honts, 2004, p. 112).

Al hilo de estos argumentos, Honts (2004) ofrece los resultados de varios estudios de campo en los que los datos del polígrafo, a diferencia de los referidos por este autor en la Tabla 22, no son examinados por evaluadores independientes, sino por evaluadores originales, es decir los mismos que aplicaron la prueba poligráfica (y que es posible que tengan información adicional del caso, conocimiento de pruebas determinantes y, en definitiva, indicios no poligráficos que pueden ser cruciales para establecer un diagnóstico). Honts (2004) justifica estos estudios alegando que reflejan mejor la realidad judicial, ya que en juicios reales, según el autor, lo habitual es que sea el propio profesional que ha pasado la prueba el que ofrezca el testimonio en el juicio. Sobre la base de estos trabajos, el autor refiere una precisión media del 96.5% para los inocentes y del 95.25% para los culpables²⁷. La conclusión final sobre la CQT, en palabras del propio autor, es la siguiente:

Las investigaciones científicas de alta calidad, tanto de campo como de laboratorio, convergen en la conclusión de que, cuando se aplica adecuadamente, la CQT es un discriminador altamente preciso de personas que mienten y dicen la verdad. Los resultados de la investigación coinciden en unas tasas de precisión de aproximadamente el 90 por cien (Honts, 2004, p. 112).

No nos extenderemos más en el debate sobre la CQT; lo expuesto hasta el momento es una muestra de lo enfrentadas que se encuentran las distintas posturas. Se dejarán para el final de este apartado las conclusiones de la NRC (2003) y Bull et al. (2004) sobre el polígrafo y las diferentes pruebas, dado que, por razones ya aludidas, estos informes se consideran más neutrales.

²⁷ Cuando Honts (2004) presenta los datos sobre la precisión de los evaluadores originales, no indica si los cálculos se hicieron sobre el total de los casos o si se excluyeron los casos no concluyentes.

La GKT (Guilty Knowledge Test)

Lykken (1998), tan defensor de la GKT como crítico de la CQT, señala que una diferencia esencial entre una persona que ha cometido un delito y otra que es inocente es que el primero sabe exactamente cómo se desarrollaron los hechos, mientras que el inocente no. Este planteamiento supone la base para la aplicación de la GKT.

En concreto, una GKT consiste en formular al sospechoso una serie de preguntas con opciones múltiples de respuesta referentes al hecho bajo investigación. Para cada pregunta, sólo una de tales opciones de respuesta debe ser cierta. Se asume que los culpables mostrarán sistemáticamente una mayor respuesta psicofisiológica ante la opción correcta que ante las demás, porque la reconocerán. Sin embargo, para los inocentes, todas las opciones serán igualmente plausibles (por ejemplo Lykken, 1998; Vrij, 2000).

Según Lykken (1998), hay dos características esenciales que debe cumplir un ítem para ser considerado bueno para esta prueba. En primer lugar, para una pregunta con cinco opciones de respuesta, la probabilidad de una persona inocente de mostrar una reacción psicofisiológica superior ante la opción correcta que ante las demás debe ser del 20%. Esto se logra si todas las alternativas se establecen de forma que parezcan igualmente plausibles para una persona que no posee “conocimiento culpable”. Conviene recordar en este punto que las alternativas no tienen porqué consistir exclusivamente en información escrita; también pueden contener fotografías, sonidos o diversos tipos de objetos. En segundo lugar, si el individuo es culpable, la probabilidad de que muestre una respuesta psicofisiológica superior ante la opción correcta que ante las demás debe ser del 80%, lo cual se consigue si la opción adecuada se presenta de forma que sea reconocible para quien tenga “conocimiento culpable”. Esta segunda condición exige controlar al individuo para que no se auto-estime en una o varias de las opciones alternativas. Otro aspecto a tener en cuenta a la hora de diseñar las alternativas de un ítem es que la primera de ellas nunca debe ser la correcta. Se trata de una alternativa

“amortiguadora”²⁸ no computable, cuyo objeto es eliminar la tendencia del individuo a reaccionar más fuertemente al primer elemento de una lista simplemente por ser el primero (Lykken, 1998).

Respecto al número de ítems que se deben incluir, teóricamente en la medida que la prueba cuente con más ítems aumentará su discriminabilidad. A este respecto, Lykken (1998) plantea lo siguiente:

En una GKT con diez ítems, para un sospechoso que fallase dos ítems las posibilidades de ser inocente serían de más de 200 frente a 1 (de ser culpable). Para un sospechoso que mostrase una respuesta más fuerte ante la alternativa correcta en nueve de los diez ítems, podemos decir que las posibilidades de que posea conocimiento culpable son mayores de 30.000 frente a 1 (de que no lo posea) (Lykken, 1998, p. 291).

Una vez que se han elaborado los ítems adecuadamente, la aplicación de la prueba es relativamente sencilla. El examinador va presentando cada ítem al examinado y, según Vrij (2000), ante cada alternativa de respuesta el examinado simplemente debe decir “no”. Lykken (1998), sin embargo, indica que es mejor que el examinado repita cada opción de respuesta, ya que esto dificulta el uso de contramedidas.

A la hora de puntuar una GKT hay varias alternativas. Lykken (1998) señala que lo más sencillo es asignar un punto por cada acierto, es decir, por cada ítem en el que la activación ante la alternativa relevante haya sido mayor que ante las demás. Según indica el autor, con diez buenos ítems una persona inocente estará probablemente por debajo de los tres aciertos, mientras que más de la mitad de los culpables deberían tener al menos siete aciertos.

Tasas de precisión obtenidas con la GKT

En la Tabla 23 se presentan los resultados de distintas revisiones sobre la precisión de la GKT, tanto en estudios de laboratorio como en estudios de campo.

²⁸ El término original empleado por el autor es *buffer*, que traducimos aquí como amortiguador.

Tabla 23. Resultados de Distintas Revisiones Sobre la Precisión de la GKT Expresados en Porcentajes Medios.

Revisiones	Mentirosos		Veraces		Global
	Correct.	Incorrect	Correct.	Incorrect.	
<i>Laboratorio</i>					
Vrij (2000)	82%	18%	96%	4%	
Lykken (1998)	88%	12%	97%	3%	93%
<i>Campo</i>					
Bull et al. (2004) ^a	42%	58%	98%	2%	
	76%	24%	94%	6%	
Vrij (2000)	59%	41%	96%	4%	

^aBull et al. (2004) solamente encontraron dos investigaciones de campo en las que se había utilizado la GKT, por lo que solamente se ofrecen los resultados de estos dos estudios.

Observando los datos de la Tabla 23 se puede apreciar que la GKT, al igual que la CQT, también parece mostrar un sesgo, pero en este caso en sentido opuesto; es decir, la GKT parece estar sesgada hacia la verdad. De esta forma, aunque las tasas de precisión para los inocentes son bastante satisfactorias (entre el 94% y el 98%), la eficacia de la prueba a la hora de clasificar correctamente a los culpables es excesivamente pobre (entre el 59% y 88%). En este caso se trata de un problema de falsos negativos, que en la práctica se traduciría en que muchos culpables conseguirían pasar la prueba y ser juzgados inocentes.

Por su parte, el informe del NRC (2003) aborda el análisis de la precisión de la GKT desde la Teoría de la Detección de Señales. La precisión se expresa como la mediana de A. El informe solamente consideró estudios de laboratorio. La capacidad estimada de la GKT para discriminar verdades de mentiras en este contexto ($A = 0.88$) fue levemente superior que la estimada para la CQT ($A = 0.85$).

Limitaciones y críticas contra la GKT

Los principales críticos de la GKT son los defensores de la CQT. Así, Honts (2004) señala varias limitaciones de esta prueba. En primer lugar, uno de los pilares de la GKT es la consideración de que nadie excepto el culpable puede saber cuál es la alternativa correcta, sin embargo, en muchas ocasiones la prensa y los medios de comunicación filtran detalles e información del caso, lo que hace que exista la posibilidad de que personas que no hayan cometido el

delito sí posean datos cruciales del mismo. Estas personas podrían correr el riesgo de no pasar la prueba aún sin haber cometido el delito. En segundo lugar, es necesario tener en cuenta que la GKT solamente evalúa el conocimiento que un individuo tiene de un suceso, por ejemplo un delito (Masip y Alonso, 2006; Vrij, 2000), no si dicho individuo ha sido el autor del mismo. En este sentido, alguien que haya sido testigo presencial del episodio puede tener el mismo conocimiento que el culpable, y sin embargo es inocente. ¿Cómo puede la GKT distinguir un testigo ocular de un delito del autor del mismo?

Otra importante limitación se relaciona con los ítems de la prueba. Según Honts (2004), nunca es posible saber con seguridad qué elementos de la escena de un crimen recordará el perpetrador. Se trata, sencillamente, de un problema de memoria, o del simple hecho de que es posible que el culpable no se fije en determinados detalles que luego pueden formar parte de los ítems de la GKT. Además, es posible que la escena del crimen haya cambiado desde que el culpable la abandonó hasta que la policía la haya podido descubrir. Por ejemplo, un delincuente puede haber herido de muerte a una persona en el dormitorio de su vivienda para robarle y luego huir. Es posible que el desafortunado agredido, antes de morir, haya intentado salir de la casa en busca de ayuda. De modo que su cadáver puede haber sido encontrado en otra parte del domicilio. Incluso es posible que otras personas hayan llegado antes a la escena del crimen y la hayan modificado, sin que la policía tenga conocimiento de ello.

Todo lo anterior hace que el número de casos en los que la GKT puede ser aplicada sea reducido (Ekman, 1985/1992; Vrij, 2000).

Resumiendo, en este apartado se han descrito las dos pruebas poligráficas más importantes, y se ha observado que ambas tienen defensores y críticos. Ambas, además, presentan importantes limitaciones. No obstante, cuando se ha pedido su opinión a la comunidad científica, gran parte de los estudiosos han mostrado una preferencia por la prueba del conocimiento culpable por tener una base científica más sólida (Masip y Alonso, 2006; Vrij, 2000), a pesar de que la prueba que se utiliza habitualmente en los juicios sobre delitos reales es la CQT.

Apunte breve sobre las contramedidas

Indefectiblemente unido al intento de utilizar el polígrafo para detectar mentiras, está sin duda el intento del culpable de engañar a la máquina. Las contramedidas constituyen intentos del examinado de alterar los resultados del polígrafo a través de la producción, intensificación, debilitamiento o supresión voluntarias de respuestas fisiológicas que lleven al examinador a concluir que está diciendo la verdad. Por ejemplo, en una CQT el examinado podría intentar aumentar la intensidad de sus respuestas fisiológicas ante una pregunta control, o disminuir la intensidad de sus respuestas ante una pregunta relevante. Lo primero podría hacerlo, por ejemplo, mordiéndose la lengua justo antes de responder o presionando con fuerza los dedos de los pies contra el suelo mientras contesta a la pregunta. La reacción ante la pregunta relevante puede reducirse mediante técnicas de distracción (contar de 3 en 3 hacia atrás empezando desde 200) o relajación (imaginarse en una apacible playa).

Honts (2004) se refiere a las contramedidas como cualquier cosa que una persona podría hacer para intentar distorsionar o alterar un examen poligráfico. Este autor clasifica las contramedidas en cuatro categorías.

a) Informacionales. Aquellas derivadas del conocimiento e información que una persona posee sobre la prueba poligráfica a la que está siendo sometido, así como de las acciones que podrían servirle para inhibir o generar respuestas fisiológicas.

b) Espontáneas. Aquellas que las personas intentan espontáneamente, es decir, sin haber recibido información ni entrenamiento de ningún tipo.

c) Generalizadas. Acciones específicas que una persona emprende con la intención de alterar su reactividad fisiológica general. En esta categoría se incluye la ingesta de drogas, el ejercicio, o ponerse sustancias anti-transpirantes en los dedos.

d) Puntuales. Aquellas contramedidas aplicadas en momentos específicos de una prueba poligráfica para intentar alterar la respuesta fisiológica concreta ante un ítem particular.

El NRC (2003) establece una clasificación de las contramedidas en tres tipos: consumo de drogas y alcohol, contramedidas físicas (si la acción específica para alterar la respuesta fisiológica es de tipo físico, por ejemplo, infringirse dolor o contraer los músculos) y mentales (si la actividad es de tipo mental, por ejemplo, contar al revés o producción de imaginación mental emocional).

Valoración del polígrafo según los informes del NRC y la BPS

Se ha dejado para el final la valoración que estas dos comisiones de profesionales han publicado sobre el polígrafo, dado que suponen las revisiones más exhaustivas y rigurosas. En la Tabla 24 se expone, a modo de esquema, un resumen de las conclusiones principales. Dado que en muchos casos las conclusiones presentadas por la BPS suponen la cita textual y ratificación de los argumentos del NRC, es de este último informe del que hemos tomado el grueso de los datos. Las ideas incluidas del informe de la BPS, se han añadido más para reflejar el acuerdo entre los dos informes que para añadir información nueva.

Tabla 24. Resumen de las Conclusiones Principales Sobre la Utilidad del Polígrafo Según los Informes del NRC (2003) y la BPS (Bull et al., 2004).

Ciencia Básica

NRC Casi un siglo de investigación en psicología y fisiología ha proporcionado una base débil respecto a la expectativa de que una prueba poligráfica podría alcanzar una precisión extremadamente alta. Las respuestas psicofisiológicas medidas con el polígrafo no se relacionan exclusivamente con el engaño (NRC, 2003, p. 212).

La fundamentación teórica del polígrafo es bastante débil, especialmente a la hora de aislar el efecto del miedo, *arousal* u otros estados emocionales que se disparan como respuesta a preguntas relevantes o control. No hemos encontrado ningún esfuerzo serio para establecer una validación de constructo de la prueba poligráfica (NRC, 2003, p. 213).

La investigación sobre el polígrafo no ha progresado a lo largo del tiempo como cualquier otro campo científico lo haría. No se ha acumulado conocimiento o fortalecido su base científica de forma significativa (NRC, 2003, p. 213).

Tabla 24 (continuación)

La inherente ambigüedad de las medidas fisiológicas que el polígrafo utiliza sugiere que un esfuerzo adicional para mejorar las técnicas poligráficas y la interpretación de los resultados, se traducirá solamente en leves mejoras en la precisión (NRC, 2003, p. 213).

BPS Aunque el equipamiento poligráfico mide con precisión una serie de actividades fisiológicas, esas actividades no reflejan un único proceso subyacente. Adicionalmente, estas actividades no son siempre necesariamente iguales ni en una misma persona ni para personas distintas (Bull et al., 2004, p 29).

NRC La evidencia referida a la validez del polígrafo se asienta principalmente en estudios empíricos sin base teórica que muestran asociaciones entre puntuaciones instantáneas derivadas de medidas poligráficas e indicadores independientes de verdad o mentira (NRC, 2003, p. 213).

La precisión puede variar como consecuencia de una serie de factores tales como la población de origen de los examinados, las características individuales tanto del examinador como del examinado, las relaciones establecidas en la entrevista, el tipo de prueba poligráfica empleada y el empleo de contramedidas (NRC, 2003, p. 213).

En buena medida, la investigación sobre la precisión del polígrafo pasa por alto aspectos críticos de las pruebas poligráficas reales. En estudios de laboratorio centrados en un incidente específico que utilizan falsos delitos, las consecuencias asociadas a mentir o ser juzgado como mentiroso casi nunca reflejan la gravedad que ello conlleva en la vida real, en las situaciones en las que se utiliza el polígrafo (NRC, 2003, p. 213-214).

Por las razones citadas, creemos que las valoraciones de la precisión del polígrafo procedentes de la investigación existente, sobreestiman la precisión en la práctica real (NRC, 2003, p. 214).

A pesar de las limitaciones respecto a la calidad de la investigación empírica y la dificultad para generalizar a situaciones del mundo real, concluimos que en poblaciones de examinados similares a las utilizadas en la literatura sobre investigación poligráfica, sin entrenamiento en contramedidas, las pruebas poligráficas centradas en un incidente específico para investigaciones sobre un caso concreto pueden discriminar la mentira de la verdad en niveles que están claramente por encima del azar, aunque claramente por debajo de la perfección (NRC, 2003, p. 214).

Tabla 24 (Continuación)

BPS La gran mayoría de los estudios publicados sobre la detección de mentiras con el polígrafo se han realizado en el laboratorio, y no han tenido mucho éxito en la recreación de muchos de los factores que podrían estar presentes en la vida real (en el “campo”), tales como los altos costes de ser juzgado como mentiroso. Por esto, persisten serios interrogantes sobre la generalización de los resultados más allá de las situaciones de laboratorio. Sin embargo, la escasez de buenos estudios de campo es un impedimento sustancial para valorar la validez científica de la detección de la mentira con el polígrafo (Bull et al., 2004, p. 29).

Respecto a los estudios de campo, el informe de la BPS concluye:

Uno de los principales problemas es establecer la verdad de base, es decir, establecer con certidumbre si el sospechoso es realmente inocente o culpable. Idealmente, esto se debería hacer con evidencia concluyente y corroborativa obtenida con independencia de la prueba poligráfica, como la evidencia del ADN. Sin embargo, normalmente se realizan pruebas poligráficas aunque este tipo de evidencia no esté disponible (Bull et al., 2004, p 14).

En los estudios de campo, las confesiones se aceptan ampliamente como una forma de establecer la verdad de base. Sin embargo, esto es problemático (Bull et al., 2004, p 29).

Contramedidas

NRC Tanto la ciencia básica como la investigación poligráfica ofrecen razones para pensar que la precisión de una prueba poligráfica puede verse reducida mediante el uso de contramedidas (NRC, 2003, p. 216).

Todos los indicadores fisiológicos medidos con el polígrafo pueden modificarse a través de un esfuerzo consciente, bien físico o cognitivo. Hay suficiente investigación empírica como para justificar nuestra preocupación de que pueden aprenderse contramedidas eficaces (NRC, 2003, p. 214).

BPS A pesar de que se han publicado pocos estudios de buena calidad sobre los efectos de las contramedidas, el uso de las mismas supone una seria amenaza, dado que las actividades fisiológicas medidas con el polígrafo podrían verse afectadas por medios físicos o cognitivos (Bull et al., 2004, p 29).

Como se puede apreciar, las conclusiones de estos dos informes, lejos de considerar al polígrafo como un buen instrumento para la detección del engaño, se muestran más bien críticas, especialmente en cuatro pilares fundamentales.

En primer lugar, se indica que la ciencia básica no garantiza la validez del polígrafo, y se alude además a la debilidad de la fundamentación teórica subyacente al mismo. En segundo lugar, aunque se admite que la precisión alcanzada con el polígrafo está por encima del nivel del azar, se deja claro que está lejos de la perfección. Al contrario de lo que se trata de hacer creer a los examinados con la prueba CQT, el polígrafo es falible. En tercer lugar, se cuestiona también la calidad de gran parte de los estudios en los que se ha utilizado este instrumento. En concreto, se critica el que muchos de ellos no hayan establecido la verdad de base de forma fiable y con independencia de la propia prueba poligráfica. Finalmente, se indica que las personas examinadas pueden alterar los resultados del polígrafo a voluntad mediante el empleo de contramedidas. A estas críticas se unen también aquellas referidas al hecho de que la utilización del polígrafo ha sido éticamente cuestionada, especialmente cuando se aplica la prueba CQT (véanse, por ejemplo, Masip y Alonso, 2006; Vrij, 2000).

En definitiva, aunque nuestra meta no es evaluar la validez del polígrafo, sí consideramos oportuno hacer referencia a dos aspectos que son incuestionables y que siempre se deberían tener en cuenta a la hora de admitir los resultados poligráficos como prueba en procesos judiciales:

a) El polígrafo no es un detector de mentiras. Tal máquina no existe. El polígrafo es sólo un aparato que mide la intensidad de distintos tipos de reacciones fisiológicas que tienen orígenes diversos, y que no se derivan exclusivamente del hecho de mentir.

b) El polígrafo es falible, no es perfecto, y dependiendo de la prueba que se utilice habrá más posibilidades de culpar erróneamente a un inocente o de que un culpable pase con éxito el examen poligráfico. Teniendo presentes estas ideas fundamentales, la prudencia debería guiar cualquier decisión sobre la admisibilidad de los resultados poligráficos, especialmente cuando el precio del error puede ser la libertad o incluso la vida de una persona.

Métodos de detección del engaño a partir de indicadores psicofisiológicos alternativos al polígrafo

Además del polígrafo, se han diseñado otros aparatos y técnicas para tratar de detectar mentiras a partir de la observación de indicadores psicofisiológicos. La mayoría de estos instrumentos parten de los mismos principios que el polígrafo, es decir, una máquina recoge distintos tipos de respuestas fisiológicas que emite una persona mientras se le hacen determinadas preguntas referidas al tema sobre el que se quiere saber si miente o no. Las señales recogidas se amplifican y se representan gráficamente, bien en la pantalla de un ordenador, bien sobre papel, de forma que sean interpretables por un profesional que debe tomar una decisión sobre la veracidad de la persona entrevistada.

Con la excepción de los analizadores de estrés vocal, las técnicas e instrumentos a los que nos referiremos a continuación se encuentran todavía en fase de experimentación y apenas hay estudios empíricos al respecto, salvo aquellos realizados por sus creadores para refinar la recogida de datos, el procedimiento de medida, o la salida de resultados. Además, en algunos casos se trata de protocolos que implican instrumental de alto coste económico o la intervención de personal muy cualificado, lo cual hace que su difusión entre el grueso de los profesionales de la detección del engaño sea, en estos momentos, escasa. En este apartado tan solo se describirán brevemente algunos de estos procedimientos. El lector interesado en una exposición más pormenorizada de cada uno de ellos puede consultar los trabajos de Ford (2006), Masip (2002), Masip y Alonso (2006), o el NRC (2003).

El analizador de estrés vocal

El analizador de estrés vocal es una máquina que permite detectar la ausencia de ciertos microtemblores que se producen en la voz, lo cual indicaría que el individuo estaría experimentando tensión (Masip et al., 2004b). Los profesionales que emplean este instrumento para la detección del engaño hipotetizan que, al hablar en condiciones normales, los músculos de la garganta y la laringe producen unos microtemblores que se pueden detectar

con el instrumento adecuado. La activación que experimenta un individuo al mentir genera una tensión que hace que dichos microtemblores desaparezcan. Este proceso sería detectable a través del analizador de estrés vocal (Lykken, 1998; Masip y Alonso, 2006; Masip et al., 2004b). A partir de esta premisa, se podría detectar cuando un individuo miente mediante la observación de las variaciones en los patrones de microtemblores que ofrecen los analizadores de tensión vocal.

Los principales promotores de este sistema de detección del engaño han ido introduciendo en el mercado, de forma paulatina, diversas versiones de analizadores de estrés vocal: el lanzamiento comercial del PSE (*Psychological Stress Evaluator*), del VSA Mark 2000 o del CVSA (*Computer Voice Stress Analyser*) constituyen una estrategia de comercialización que ha ido alzando los precios del producto a medida que se presentaba una nueva versión. Pero ¿funcionan realmente los analizadores de estrés vocal? No nos detendremos en esta Tesis a analizar aquellos estudios (algunos de ellos no publicados) realizados por profesionales de algún modo afines a las empresas que comercializan estos aparatos y que, tal como se indica en algunas revisiones (Lykken, 1998; Masip et al., 2004b), pueden ofrecer resultados que quizá estén sesgados. Bastará para este trabajo citar las conclusiones de dos revisiones independientes cuyo objetivo ha sido arrojar luz sobre la fiabilidad y la validez de un instrumento que, tal como indican Masip et al. (2004b), si funcionase ofrecería múltiples ventajas, tales como poder tomar datos sin que el individuo lo advirtiera, poder ser utilizado con cintas grabadas, un entrenamiento sencillo y el poco coste del aparato en comparación con instrumentos como el polígrafo. El problema es que no funciona. Lykken (1998), tras revisar distintos estudios sobre a los analizadores de la tensión vocal, concluye:

No hay evidencia científicamente creíble de que el PSE, el CVSA, el VSA Mark 1000, el Hagoth, el Truth Phone, o cualquier otro instrumento disponible pueda medir con fiabilidad diferencias en estrés que se reflejen en la voz humana. Hay considerable evidencia de que esos aparatos, utilizados conjuntamente con interrogatorios estandarizados para la detección de la mentira, discriminan el engaño de la verdad con un nivel de precisión aproximado al azar (Lykken, 1998, p. 112-113).

Posteriormente, Masip et al. (2004b) publican una revisión de los resultados de distintos estudios sobre analizadores de la tensión de la voz. Las conclusiones son claras. Para empezar, según estos autores, la investigación cuestiona no solamente que se produzcan cambios en la tensión de la voz cuando el hablante experimenta estrés, sino la propia existencia de los microtemblores vocales. Tampoco se puede probar, a partir de los resultados de los distintos estudios científicos revisados, que el espectro de la voz varíe en función de si un individuo miente o dice la verdad. Estas dos conclusiones bastarían para invalidar la hipótesis que sustenta la utilidad de los analizadores de estrés vocal. No obstante, Masip et al. (2004b) añaden dos importantes y definitivas conclusiones: (a) la investigación no ha conseguido probar de forma clara y contundente que los distintos aparatos comercializados como analizadores de tensión vocal realmente midan la tensión de la voz y no el efecto de otras variables extrañas; y (b) los resultados de distintos estudios muestran que la precisión de estos instrumentos a la hora de detectar el engaño no supera los niveles del azar.

En definitiva, la investigación científica ha mostrado que esta aproximación a la detección del engaño es claramente ineficaz, por lo que nuestra postura no puede ser sino contraria al empleo de los distintos analizadores de estrés vocal que actualmente se comercializan. Ningún aparato que no haya demostrado previamente su fiabilidad y validez a nivel científico para discriminar verdades de mentiras debería utilizarse a tal efecto.

Los Potenciales Evocados y la Huella Dactilar Cerebral

Cuando a una persona se le presenta un estímulo, la percepción del mismo va a generar en su cerebro una serie de ondas o picos que variarán en sus características formales en función del tipo y caracteres del estímulo presentado. A estos picos se les denomina Potenciales Evocados (PE), y son susceptibles de observación y estudio a través de un electroencefalograma (EEG). El trabajo con los PE como herramienta para la detección del engaño se basa en la observación de la forma de esas ondas o picos. En este caso, la onda a tener en cuenta se denomina P300. Este pico fue inicialmente descrito por Sutton (1965, citado en Ford, 1998). La P300 aparece reflejada en el EEG

cuando una persona se somete a una serie de estímulos y uno de ellos le resulta especialmente llamativo o significativo (Rosenfeld, 2002). La P300 se ha empleado para detectar el engaño utilizando la prueba del conocimiento culpable (GKT) a la que nos hemos referido anteriormente. Por ejemplo, Rosenfeld (2002) describe como, en una serie de distintos estudios realizados junto con su equipo en los años 80, observaron que ante el ítem significativo de cada pregunta de la GKT, la P300 recogida por un EEG mostraba características determinadas en cuanto a su tamaño y amplitud que no aparecían ante los demás ítems. De esta forma, se ha tratado de describir un perfil específico de la P300 en aquellos casos en los que el individuo miente, y otro distinto cuando dice la verdad (Rosenfeld, 2002).

Siguiendo estos planteamientos, Lawrence Farwell ha realizado una serie de estudios en los que ha examinado la denominada “Respuesta Encefalográfica Multifacética”, es decir, las diferentes características de las distintas ondas que aparecen en el EEG de un individuo sometido a una GKT. Farwell ha concluido la existencia de una *Brain Fingerprint*, o huella dactilar cerebral específica, que ha denominado MERMER (*Memory and Encoding Related Multifaceted Electroencephalographic Response*), que aparece cuando el individuo reconoce información especialmente relevante para él. Dicha huella puede observarse con el equipamiento necesario para realizar un MERA (*Multifaceted Electroencephalographic Response Analysis*), es decir, un análisis de la respuesta electroencefalográfica multifacética. Este análisis consiste básicamente en la observación y comparación de distintos parámetros de los potenciales evocados (aislándolos de otro tipo de actividad cerebral) que aparecen cuando a un individuo se le presenta un estímulo determinado (véase, por ejemplo, Farwell y Smith, 2001, para una explicación más detallada de cómo se ha desarrollado y en qué consiste este método de análisis). Farwell y Smith (2001) indican haber realizado un total de cinco estudios en los que han obtenido unas tasas de precisión cercanas al 100% en la detección de información oculta basándose en si aparecía o no un MERMER en la respuesta de los participantes ante distintos tipos de estímulos.

En cualquier caso, antes de caer en un optimismo excesivo es necesaria más investigación, no solamente respecto a la eficacia de la *Brain Fingerprint*,

sino de los potenciales evocados en general y de la onda P300 en particular. Sin duda, esta forma de detectar el engaño no está carente de problemas. Uno de ellos es el uso de contramedidas. Rosenfeld, Soskins, Bosh y Ryan (2004) concluyeron que hay contramedidas muy fáciles de aprender que resultan muy eficaces en una prueba basada en la P300. Además, cuando tras varias semanas repitieron dicha prueba con los mismos participantes, pidiéndoles en este caso que no las utilizaran, el hecho de haberlas empleado en la prueba anterior seguía afectando a la forma de la P300, de manera que algunos participantes culpables no fueron detectados. Si bien los autores indican que podría identificarse a un individuo que estuviera empleando contramedidas a partir de sus tiempos de reacción a los estímulos presentados, también manifiestan su preocupación por el hecho de que, ante una persona bien entrenada para mentir y utilizar este tipo de estrategias, es posible que ni siquiera la observación de sus tiempos de reacción fuera útil para clasificarlo correctamente como mendaz.

Otra importante limitación de esta técnica es que, tal como Farwell y Smith (2001) indican, es necesario que la persona a la que se le administra la prueba muestre un talante colaborador. Es decir, debe seguir las instrucciones que le indique el profesional sobre aspectos como permanecer en la misma postura y no moverse, mantener la mirada fija o concentrarse solamente en la tarea. De lo contrario, el nivel de interferencias podría hacer inviable la interpretación de los datos. Si bien en un estudio científico se espera que los participantes realmente colaboren con el investigador, quizá no se pueda decir lo mismo en un caso real, en que el individuo interrogado puede tener “especial interés” en que la prueba sea inconcluyente. A estas limitaciones hay que añadir también las propias de la GKT (descritas en el apartado correspondiente), que se ha empleado en la mayoría de los casos. Finalmente, al margen de todos estos posibles problemas, conviene recordar que en la mayoría de los casos, tal como se practica este tipo de pruebas, lo más que pueden probar es que un individuo tiene un determinado conocimiento, no que ha cometido un delito.

La Resonancia Magnética Funcional (fMRI)

Este es un método relativamente nuevo para la detección del engaño. A partir de los datos ofrecidos por una resonancia magnética funcional, se ha tratado de observar qué zonas del encéfalo se activan cuando una persona miente (Phan et al., 2005). Para ello se ha empleado la prueba del conocimiento oculto (GKT).

Utilizando este sistema, distintos estudios (por ejemplo, Davatzikos et al., 2005; Kozel et al., 2005; Phan et al., 2005) han obtenido, en situaciones muy controladas de laboratorio, unas tasas de precisión en la detección de mentiras de entre el 88% y el 99%. Aunque estos resultados son llamativos y esperanzadores, lo cierto es que la detección del engaño empleando la *fMRI* tiene limitaciones muy importantes.

En primer lugar, tal como señala el informe del NRC (2003), no se han encontrado áreas cerebrales que se activen exclusivamente cuando una persona miente. La investigación ha mostrado que determinadas zonas, como por ejemplo, áreas del cortex prefrontal, se activan al mentir (Davatzikos et al., 2005; Kozel et al., 2005; Langleben et al., 2002; Lee et al., 2002; Phan et al., 2005; Spence et al., 2001), pero éstas también se activan al realizar otros procesos más básicos, como la inhibición de la respuesta o la modificación de la conducta, que subyacen al acto de mentir (por ejemplo, Bush et al., 1998; Elliot, Dolan y Frith, 2000; MacDonald, Cohen, Stenger y Carter, 2000). Además del cortex frontal y prefrontal, cuya activación al mentir se ha observado en la mayoría de los estudios, se han identificado también otras áreas implicadas en el engaño, como la corteza cingular anterior (por ejemplo, Ganis, Kosslyn, Stose, Thompson y Yurgelun-Todd, 2003; Kozel et al., 2005) y la zona orbitofrontal derecha (Kozel et al., 2005). No obstante, estas regiones también están involucradas, por ejemplo, en la memoria episódica y la toma de decisiones. Langleben et al. (2002) encontraron que mentir activa zonas de la corteza frontal superior, pero éstas a su vez están implicadas en otras tareas cognitivas, como la memoria de trabajo, la planificación, la iniciación de tareas y el control cognitivo. Phan et al. (2005) observaron la activación del surco temporal superior cuando se miente, área que a su vez participa en los juicios

sociales, la evaluación de la intención, la confianza o la cooperación. En definitiva, y para no extender en exceso este apartado, señalaremos simplemente que la evidencia muestra que las áreas que se activan al mentir están también relacionadas con otros procesos, como el control cognitivo, la memoria y la inhibición de respuestas.

No obstante, es lógico que al mentir se activen zonas cerebrales relacionadas con funciones como la toma de decisiones, la memoria de trabajo, la planificación, la iniciación de tareas, el control cognitivo, la conciencia de las propias emociones o la inhibición de respuestas, ya que se trata de procesos cognitivos presentes en muchos tipos de mentiras. El problema es que dichos procesos no son exclusivos de la mentira, y podrían darse cuando una persona está siendo sincera. De esta forma, es posible observar cómo las mismas zonas cerebrales que se activan al mentir pueden hacerlo también al decir la verdad.

En segundo lugar, Masip y Alonso (2006) plantean que la investigación se ha desarrollado en el laboratorio, es decir, en un entorno muy controlado. Quizá en situaciones reales en las que las mentiras pueden ser más complejas y el número de estímulos y variables que afectan a la actividad cerebral no esté tan controlado, el sistema pierda su eficacia. Hay que tener en cuenta que en cualquier actividad, física o cognitiva, participa el cerebro. Durante un interrogatorio en una situación real, no sólo hay que reconocer un ítem (GKT) o presionar el botón “sí” o el botón “no” para responder. Es necesario seleccionar palabras y articularlas para que el discurso sea coherente. Hay que manejar las emociones, en algunos casos asociadas a hechos o recuerdos relacionados con el interrogatorio. En definitiva, cualquier “cosa” que el interrogado perciba, desde la iluminación hasta el color de la habitación o incluso la comodidad de la silla en la que está sentado, va a reflejarse de alguna forma en su actividad cerebral. A todo esto hay que añadir el hecho de que la investigación se ha desarrollado en un contexto en el que la persona que es interrogada ha decidido voluntariamente colaborar con el interrogador, cosa quizá poco probable en casos reales. ¿Qué pasaría si en un interrogatorio real el sospechoso decide pensar simultáneamente en otros episodios de su vida, hacer operaciones de cálculo mental, o realizar cualquier otra operación que

requiera algún esfuerzo cognitivo o tenga implicaciones emocionales? Ni siquiera cuando una persona está dispuesta a colaborar se puede hacer que su cerebro deje de procesar todos los pensamientos, emociones, sensaciones, etc. que experimenta de forma simultánea e ininterrumpida.

Finalmente, tal como Masip y Alonso (2006) señalan, se trata de un método que requiere un instrumental muy costoso y la presencia de profesionales muy cualificados. Todo lo dicho hace que en casos reales de mentiras complejas no pueda aplicarse este procedimiento.

Análisis de Imágenes Termográficas Faciales (AITF)

La investigación que utiliza el Análisis de Imágenes Termográficas Faciales (AITF) como herramienta para detectar el engaño parte de la premisa de que, por razones evolutivas de adaptación al medio, en situaciones en las que una persona experimenta ansiedad se produce un aumento del aporte sanguíneo a determinadas partes de la anatomía humana (Pavlidis y Levine, 2002a). Una de tales zonas es la periorbital ocular. Con el aparataje adecuado, es posible observar cómo dicho aporte extraordinario de sangre se traduce en un aumento de la temperatura de la zona. Asumiendo que un individuo que miente experimenta estrés, por ejemplo al considerar la posibilidad de algún tipo de castigo si su mentira es descubierta, es posible utilizar el AITF para detectar mentiras.

Pavlidis y Levine (2002b) realizaron un estudio en el que utilizaron un contexto experimental basado en un falso crimen para probar la utilidad de las AITF. A algunos de los participantes se les dio instrucciones para que actuaran como supuestos culpables, mientras que otros desempeñaron el rol de inocentes. Posteriormente, todos los participantes fueron sometidos a una prueba poligráfica CQT y otra GKT en las que negaron haber cometido el crimen. Se obtuvieron datos por los canales poligráficos habituales, pero además se tomaron imágenes termofaciales mientras los participantes contestaban a las preguntas durante el interrogatorio. Con las medidas poligráficas tradicionales se alcanzó una precisión del 63% para las mentiras y del 80% para las verdades. Para emitir los juicios de veracidad basándose en

el AITF se desarrolló un algoritmo que determinó el valor de un umbral numérico. Las imágenes termofaciales que mostraran un valor por encima de dicho umbral se interpretarían como signo de engaño, mientras que si el valor se situaba por debajo se concluiría que el participante decía la verdad. Examinando las imágenes termofaciales obtenidas ante las respuesta a la pregunta “¿Lleva usted ahora mismo encima el billete robado de 20 dólares?”, la precisión alcanzada con esta metodología fue del 75% para las mentiras y del 90% para las verdades²⁹ (Pavlidis y Levine, 2002b; véase también Pavlidis, Eberhard y Levine, 2002). No obstante, Pavlidis y Levine (2002b) no hacen referencia al AITF como un método alternativo a la prueba poligráfica, sino más bien como un canal poligráfico adicional de información que se podría integrar con el resto de canales tradicionales.

Una limitación de la termografía facial consiste en la necesidad de que la persona bajo examen permanezca lo más inmóvil posible para así poder obtener imágenes con la suficiente calidad como para poder ser analizadas. Al final de su estudio, los propios Pavlidis y Levine (2002b) plantean la necesidad de refinar el procedimiento para evitar distorsiones generadas por movimientos faciales. Tales distorsiones pueden producir interferencias que hacen muy difícil la interpretación de los datos.

Tsiamyrtzis et al. (2007) llevaron a cabo una investigación en la que se dio un importantísimo paso a delante para solucionar este tipo de problemas. Los autores desarrollaron un sistema de rastreo tándem de termo-imágenes faciales en el que dos cuadros de enfoque recogen información de forma simultánea de forma que los datos obtenidos por uno de los enfoques sirven para guiar y, en su caso, corregir posibles deficiencias del otro en caso de que la persona que está siendo entrevistada mueva el rostro (véase Tsiamyrtzis et al., 2007 para una explicación más detallada). Otro aspecto importante de este estudio es que para depurar la información obtenida, los investigadores

²⁹ Es necesario considerar estos resultados con cierta precaución, dadas las limitaciones de este estudio (por ejemplo, de las posibles comparaciones obtenidas de las dos pruebas poligráficas administradas, sólo se pudo examinar una pregunta), así como las deficiencias metodológicas que los propios autores señalan. De hecho, el informe del NRC (2003) considera que los datos obtenidos en esta investigación carecen de valor científico aceptable.

recurrieron a un sistema de eliminación de ruido de fondo en lugar de la tradicional amplificación de la señal. En varias pruebas preliminares, los autores comprobaron que esta nueva metodología era menos sensible tanto al ruido como a las interferencias producidas por movimientos faciales, obteniéndose finalmente unos datos gráficos mucho más precisos que en estudios anteriores en los que se había trabajado con un solo enfoque y se había amplificado la señal.

Tsiamyrtzis et al. (2007) emplearon esta modalidad de obtención de termogramas faciales en un estudio en el que un grupo de participantes debía mentir o decir la verdad sobre el robo de un cheque. Se alcanzó una tasa de precisión del 87.5% para las mentiras y del 86.7% para las verdades (Tsiamyrtzis et al., 2007).

El análisis de imágenes termográficas faciales ofrece varias ventajas respecto a procedimientos anteriores. Quizá una de las más interesantes es que, tal como se indica en Pavlidis y Levine (2002b), a diferencia de la prueba poligráfica tradicional no se trata de un método invasivo. Además, Masip y Alonso (2006) señalan distintas ventajas respecto a otros métodos psicofisiológicos, como el hecho de que el coste del equipo necesario sea sensiblemente inferior (a diferencia, por ejemplo, de la Resonancia Magnética Funcional), siendo también menor el tiempo requerido para la realización de la prueba y el análisis de los datos. Tampoco es necesario que el profesional que realiza la prueba posea un nivel de especialización tan alto como el necesario para operar con otros métodos.

No obstante, aunque el trabajo de Tsiamyrtzis et al. (2007) muestra un gran avance a la hora de refinar el procedimiento de recogida de información y de depuración de los datos obtenidos, es necesario seguir trabajando en esta línea. Futuras investigaciones deberían empezar a valorar el efecto de variables que se controlan en el laboratorio pero no en situaciones reales: haber comido recientemente, el maquillaje, el estado de fatiga, o si se ha realizado ejercicio físico, ya que éstas son variables que pueden afectar a los resultados (Masip y Alonso, 2006).

Los policías como detectores profesionales de la mentira y el Investigator bias effect

5

Anteriormente se han presentado distintos enfoques o aproximaciones para la detección de mentiras. Al hacerlo, se han descrito estudios que, en la mayoría de los casos, han utilizado como detectores a muestras de estudiantes que participaban en algún taller o curso extraordinario, o bien a personas especializadas en la recogida e interpretación de los datos de los distintos instrumentos empleados.

En este apartado se examina la habilidad en la detección del engaño (tasas de precisión) de aquellas personas para quienes, por su labor profesional, es importante ser capaz de distinguir cuando alguien está mintiendo y cuando está diciendo la verdad. Dentro de este grupo, al que nos referiremos en ocasiones como “profesionales” de la detección de mentiras, se podría englobar a jueces, abogados, guardias civiles, psicólogos clínicos, agentes de aduanas, inmigración o inteligencia, y otras personas que desempeñan labores similares. Sin embargo, aquí nos centraremos especialmente en los policías.

Resulta evidente que reconocer con exactitud cuando alguien está mintiendo o diciendo la verdad es una importante habilidad policial (sobre las funciones del Cuerpo Nacional de Policía véanse, por ejemplo, Abrunhosa, Soeiro y Ventura, 2006; Alvarez, 2000; Clemente, 1995; Fernández, 1990; Fernández, 2003; Rico, 1988). Pero ¿realmente la precisión de los policías al diferenciar entre verdades y mentiras se corresponde con la que cabría esperar de ellos dado su trabajo? En las próximas páginas intentaremos responder a esta cuestión mediante el examen de la literatura científica relevante.

LA PRECISIÓN DE LOS POLICÍAS

RESULTADOS GENERALES SOBRE PRECISIÓN

Distintos estudios han encontrado que (quizá sorprendentemente) los policías, a la hora de distinguir verdades de mentiras, no muestran una

precisión superior a la obtenida por observadores legos³⁰, como por ejemplo estudiantes universitarios.

En la Tabla 25 ofrecemos los resultados de algunos trabajos al respecto. Se incluyen solamente los datos referidos a aquellos grupos de participantes que no recibieron ningún entrenamiento. En un apartado posterior de esta Tesis se observará el efecto del entrenamiento en la detección del engaño sobre la precisión y los juicios de veracidad de los policías.

Tabla 25. *Porcentajes de Precisión en la Detección del Engaño.*

Estudio	Participantes	N	Precisión		
			Verdades	Mentiras	Global
Akehurst et al. (2004) ^a	Policías	26			68
	Trabajadores sociales	14			75
	Estudiantes	18			57
Bond (2007) Prueba 1	Agentes de la ley	112	59	45	52
	Estudiantes	122	66	39	52
Prueba 2	Agentes expertos	8	55	63	59
	Estudiantes	8	63	33	48
DePaulo y Pfeiffer (1986)	Policías veteranos	144			52
	Policías noveles	114			53
	Estudiantes	161			54
Ekman y O'Sullivan (1991)	Jueces	110			57
	Poligrafistas federales	60			56
	Policías	126			56
	Agentes Servicio Secreto	34			64
	Psiquiatras	67			58
	Personas interesadas	73			55
	Estudiantes	39			53
Ekman, O'Sullivan y Frank (1999)	Agentes de la CIA	23			73
	Sheriffs	43			66
	Policías	36			51
	Jueces	84			62
	Psicólogos clínicos interesados ^b	107			68
	Psicólogos clínicos sin interés ^c	209			62
	Psicólogos académicos ^d	125			58
Elaad (2003)	Policías	60	48	51	49
Garrido et al. (2004)	Policías	121	26	69	47
	Estudiantes	146	51	66	59

³⁰ Nos referimos a la precisión de personas no profesionales (estudiantes en la mayoría de los casos) al juzgar la veracidad de declaraciones sin ninguna herramienta, instrucción específica o información sobre qué indicios utilizar, basándose simplemente en su intuición y sus creencias.

Tabla 25 (Continuación)

Estudio	Participantes	N	Precisión		
			Verdades	Mentiras	Global
Hartwig, Granhag, Strömwall y Kronkvist (2006) ^e	Policías	41	57	55	56
Hartwig, Granhag, Strömwall y Vrij (2004)	Policías	30			57
Kassin, Meissner y Norwick (2005)					
Condición vídeo	Detectives de policía	28			42
	Estudiantes	29			53
Condición audio	Detectives de policía	29			55
	Estudiantes	32			64
Köhnken (1987) ^f	Policías	20	63	32	47
Leach, Talwar, Lee, Bala y Lindsay (2004) ^g					
Experimento 1	Policías	25			44
	Agentes de aduanas	48			49
	Estudiantes	32			51
Experimento 2	Policías	28			63
	Agentes de aduanas	35			62
	Estudiantes	40			64
Experimento 3	Policías	47			59
	Agentes de aduanas	40			57
	Estudiantes	39			61
Mann y Vrij (2006)	Policías	27	70	66	68
Mann, Vrij y Bull (2004)	Policías	99	64	66	65
Mann, Vrij, Fisher y Robinson (2007)					
Visual	Policías	37	34	53	43
Audio	Policías	31	63	69	66
Visual + audio	Policías	35	60	70	65
Masip et al. (2003c)	Policías	224	30	74	52
Meissner y Kassin (2002)	Policías	44			50
	Estudiantes	20			56
Porter, Woodworth y Birth (2000) ^h	Policías	32	20	60	40
	Estudiantes	32	42	42	42
Vrij (1992)	Detectives de policía	91	51	46	49
Vrij (1994)	Detectives de policía ⁱ				50
	Detectives de policía ^j				47
	Detectives de policía ^k				51
	Detectives de policía ^l				44

Tabla 25 (Continuación)

Estudio	Participantes	N	Precisión		
			Verdades	Mentiras	Global
Vrij, Akehurst, Brown y Mann (2006)					
Emisores niños (5 a 6 años)	Profesores	36	} 61	63	62
	Trabajadores sociales	37			
	Policías	37			
	Estudiantes	40			
Emisores adolescentes (14 a 15 años)	Profesores	36	} 67	57	62
	Trabajadores sociales	37			
	Policías	37			
	Estudiantes	40			
Emisores adultos	Profesores	36	} 58	63	60
	Trabajadores sociales	37			
	Policías	37			
	Estudiantes	40			
Vrij y Mann (2001a)	Policías	65	70	57	64
Vrij y Mann (2001b)	Policías	52			51
Vrij et al. (2008)					
Orden cronológico	Policías	24	50	42	46
Orden inverso	Policías	31	56	60	58
Vrij, Mann, Kristen y Fisher (2007)					
Entrevista facilitadora	Policías	26			52
Entrevista BAI	Policías	18			51
Entrevista acusatoria	Policías	24			48
Vrij, Mann, Robbins y Robinson (2006) ^m	Policías	37	70	73	72
Vrij y Winkel (1993)	Detectives de policía	91	51	46	49

^a Resultados de la fase de pre-entrenamiento.

^b Psicólogos clínicos muy interesados en la detección del engaño.

^c Psicólogos clínicos sin un interés especial en la detección del engaño.

^d Psicólogos académicos con distintas orientaciones profesionales excepto la clínica.

^e Resultados del grupo sin entrenamiento.

^f Resultados del grupo control.

^g Los emisores fueron niños.

^h Resultados de los grupos sin entrenamiento.

ⁱ Los detectores veían declaraciones que no habían sido previamente planificadas.

^j Los detectores veían declaraciones habían sido previamente planificadas.

^k Los detectores veían dos declaraciones (una verdadera y otra falsa) en las que se mostraba todo el cuerpo del emisor.

^l Los detectores veían dos declaraciones (una verdadera y otra falsa) en las que se mostraban solo las manos del emisor y el sonido se había eliminado.

^m El objetivo de este estudio era medir la consistencia en la precisión. Para ello, cada uno de los 37 participantes hizo cuatro pruebas similares de detección del engaño. Los datos que se presentan son el promedio de las cuatro pruebas.

En la Tabla 25 se presentan los datos de 27 estudios que incluyen un total de 46 muestras en las que los participantes eran policías y 3 muestras

compuestas por agentes de la ley (agentes de la CIA, agentes del Servicio Secreto y *Sheriffs*) que por sus características específicas no los consideraremos conjuntamente con los policías³¹. El resto de muestras están formadas por otros profesionales y por grupos de estudiantes que distintas investigaciones han incluido para establecer comparaciones entre la precisión de profesionales y no profesionales.

Precisión global

Aún siendo conscientes de que las condiciones experimentales y los tamaños muestrales de cada estudio son diferentes, si se promedian las tasas de precisión global de las 46 muestras de policías que aparecen en la Tabla 25 se obtiene una precisión media del 54.33, lo cual no dice mucho a favor de la habilidad de los profesionales. De hecho, contrasta con las estimaciones que los propios policías hacen de que pueden diferenciar verdades y mentiras con una precisión del 77%, tal como señala una encuesta reciente que Kassin et al. (2007) han hecho a 631 policías estadounidenses y canadienses. Adicionalmente, en la Tabla 25 aparecen 16 casos o condiciones de distintos estudios en los que se incluyen muestras de policías y estudiantes³². Comparando ambos grupos, se observa que la precisión para los policías ($M = 54.57$) fue algo inferior que la de los estudiantes ($M = 56.13$).

En general, estos datos reflejan los de dos recientes meta-análisis en los que se ha comparado la precisión de profesionales y no profesionales. El trabajo de Aamodt y Custer (2006) examinó las tasas de precisión encontradas en distintos estudios que incluían grupos de policías, detectives, agentes aduaneros, agentes federales, agentes del Servicio Secreto y agentes encargados del control de la libertad condicional de reclusos, y la comparó con

³¹ No consideraremos a estos agentes de la ley conjuntamente con los policías (excepto en el estudio de Bond, 2007, ya que al autor no indica cuantos profesionales había de cada tipo), en primer lugar porque el interés de esta tesis se centra en policías y no en agentes federales, poligrafistas o agentes de distintas agencias de inteligencia, y en segundo lugar, porque estos profesionales quizá hayan tenido un tipo de formación y experiencia laboral especial diferente del resto de los policías, dadas las características de su trabajo.

³² En el estudio de Vrij, Akehurst, et al. (2006), los autores indican que no hubo diferencias significativas entre la precisión de los profesores, trabajadores sociales, policías y estudiantes, y ofrecen los datos de estos cuatro grupos conjuntamente. Así pues, para el cálculo de la precisión de los trabajos expuestos en la Tabla 25, se atribuyó a los policías en cada condición del estudio de Vrij, Akehurst, et al. (2006) la precisión global que estos autores ofrecen.

la precisión de un grupo de estudiantes. En la Tabla 26 se muestran los resultados para cada grupo.

Tabla 26. *Porcentajes Medios de Precisión de Grupos de Profesionales (Aamodt y Custer, 2006).*

Grupo	Número de estudios	Número de participantes	Porcentaje de precisión
Policías	12	655	55.30
Detectives	7	758	50.80
Agentes de aduanas	3	123	55.30
Agentes federales	4	341	54.54
Agentes del Servicio Secreto	1	34	64.12
Agentes de libertad condicional	1	32	40.42
Estudiantes	156	11647	54.22

Como se puede observar, la precisión hallada por Aamodt y Custer (2006) para los policías ($M = 55.30$) es similar a la observada en la revisión realizada en esta Tesis ($M = 54.33$). Al comparar la precisión de los policías con la de los estudiantes ($M = 54.22$), Aamodt y Custer concluyen que no hay diferencias³³.

En el meta-análisis de Bond y DePaulo (2006) al que ya nos hemos referido en un apartado anterior (ver el epígrafe sobre la precisión en la detección de mentiras del apartado referido a la aproximación basada en los indicios no verbales y paralingüísticos), se comparó la precisión obtenida con “expertos” de la detección del engaño (policías, agentes de aduanas, detectives, agentes federales, poligrafistas profesionales, jueces, especialistas interrogadores, auditores, expertos en entrevistas de trabajo y delincuentes) con la precisión obtenida con personas sin ningún interés especial por la detección del engaño. El porcentaje de precisión obtenido para los “expertos” fue del 53.81%. La precisión para los no expertos fue del 53.29%. Entre los dos grupos no hubo diferencias significativas.

Si la precisión obtenida a partir de los estudios de la Tabla 25 es ya de por sí poco satisfactoria, ésta podría estar incluso sobre-estimada, dados los atípicos resultados de algunos estudios, debidos quizá a cuestiones metodológicas. En

³³ No obstante, se aprecia que los agentes del Servicio Secreto muestran unas tasas de precisión superiores al resto de los grupos. En este caso, los resultados se obtuvieron solamente del estudio de Ekman y O’Sullivan (1991). Aunque esta Tesis se centra en policías y no en agentes de inteligencia, de la CIA o de distintos servicios secretos, al final de este apartado se hará una breve mención a estos grupos y los estudios en los que han sido incluidos.

concreto, los resultados de algunos trabajos del grupo de Aldert Vrij (Akehurst et al., 2004; Mann y Vrij, 2006; Mann et al., 2004; Mann et al., 2007; Vrij, Akehurst, et al., 2006; Vrij, Mann, Robbins, et al., 2006), sobresalen por haber encontrado unas tasas de precisión muy elevadas ($M = 65.47$ para las muestras obtenidas de estos estudios tomadas conjuntamente). A continuación se revisarán con más detalle estos estudios.

En el estudio de Akehurst et al. (2004) se contó con un total de 26 policías que realizaron dos tareas de detección del engaño: una sin entrenamiento, y posteriormente otra tras haber recibido cierto entrenamiento. Nos centraremos aquí en la tarea pre-entrenamiento. Hay algunos detalles que podrían haber influido en los resultados de esta tarea. En primer lugar, cada uno de los 26 policías (un tamaño muestral reducido) juzgó únicamente 4 declaraciones, de forma que solamente tenían que emitir cuatro juicios de veracidad. Cuando el número de participantes y declaraciones a juzgar es tan reducido, la posibilidad de obtener resultados por azar aumenta significativamente. Frank y Feeley (2003) recogen la opinión de la gran mayoría de investigadores de la detección del engaño cuando aconsejan que, en las pruebas de detección, el número de declaraciones a juzgar esté en torno a 10.

Adicionalmente, a cada policía se le presentaron declaraciones elegidas al azar, con lo que en algunos casos dos eran verdad y otras dos mentira, pero en otros todas eran verdad o todas eran mentira, o había tres verdades y una mentira, o tres mentiras y una verdad. Aunque podríamos esperar que unas combinaciones contrabalanceasen otras, hay que tener en cuenta que un policía que tuviese un sesgo hacia la mentira (véase Meissner y Kassin, 2002) e hiciese 4 juicios de mentira, podría obtener una precisión del 100% si debía juzgar cuatro mentiras, o del 0% si debía juzgar 4 verdades. Las puntuaciones no reflejarían la precisión real, sino el sesgo de los policías. En ambos casos estaríamos ante puntuaciones extremas que, teniendo una muestra de tan solo 26 participantes, podrían marcar diferencias importantes en uno u otro sentido al calcular la precisión media.

Por otra parte, la precisión del 57% de los estudiantes es similar a la que se ha encontrado en otros estudios; sin embargo, una precisión del 75% como la

de los trabajadores sociales se sitúa muy por encima de los resultados habitualmente encontrados. El hecho de que estas puntuaciones fuesen muy diferentes y sin embargo no se encontraran diferencias significativas entre los tres grupos, puede ser un indicio de que hay que tomar los datos con precaución.

Respecto a los trabajos de Mann y Vrij (2006), Mann et al. (2004), Mann et al. (2007) y Vrij, Mann, Robbins, et al. (2006), en todos ellos los investigadores partieron del mismo conjunto inicial de declaraciones para elaborar los vídeos, las cuales tenían importantes limitaciones. Más en concreto, Mann et al. (2004) y Vrij, Mann, Robbins, et al. (2006) utilizaron las mismas 54 declaraciones. La longitud de las mismas varió ostensiblemente (6 segundos la más corta y 145 la más larga). Si bien no es recomendable la utilización de declaraciones excesivamente cortas, ya que pueden carecer de elementos en los que basarse para emitir un juicio, es aún menos deseable utilizar declaraciones de muy distinta duración si éstas se van a comparar después, ya que esta variable (duración) podría tener un efecto sobre los juicios.

Las diferencias en los tiempos propiciaron además diferencias en el número de declaraciones en cada uno de los vídeos (15, 16, 10 y 13, respectivamente). La fatiga pudo afectar de forma diferencial a quienes juzgaron 10 declaraciones y a quienes juzgaron 16. Tampoco estuvo equilibrado el número de veces que cada declarante podía aparecer (un mínimo de 2 veces y un máximo de 8). Cuando un declarante aparece en más de una ocasión se corre el peligro de no juzgar la veracidad de su declaración en sí misma, sino por comparación de su conducta con la mostrada en otras declaraciones.

Además, en el estudio de Vrij, Mann, Robbins, et al. (2006), los propios autores ofrecen datos que hacen pensar que algunos de los *clips* eran mucho más transparentes (es decir, la veracidad real era más fácil de reconocer) que otros. De esta forma, se encontró que la precisión para algunas declaraciones se situó entre el 14% y el 45% mientras que para otras se ubicó entre el 91% y el 100%. En concreto, la precisión más baja para un *clip* fue del 14%, y la segunda más baja fue del 30%. Para cinco *clips* se obtuvo una precisión de entre el 31% y el 45%, para otros cuatro de entre el 46% y el 55%, para otros

10 la precisión se situó entre el 56% y el 70%, para otros nueve entre el 71% y el 80%, para otros 11 entre el 81% y el 90% y para otros 13 entre el 91% y el 100% (Vrij, Mann, Robbins, et al., 2006). Así pues, el hecho de que muchas declaraciones fueran excesivamente “transparentes” respecto a la veracidad real pudo “inflar” la precisión final encontrada.

En este último estudio, en que el objetivo general era observar la consistencia en la precisión de un grupo de policías a través de cuatro tareas de evaluación de la credibilidad (en cada una de las tareas se utilizó uno de los cuatro vídeos creados) hay que mencionar una importante limitación adicional: los autores no informan de si los participantes tuvieron la oportunidad de hablar entre ellos en los distintos periodos de tiempo entre las cuatro pruebas. Si éste fuera el caso, cabría la posibilidad de que los juicios de unos se vieran influidos por los de los otros, especialmente teniendo en cuenta que un mismo declarante podía aparecer en más de un vídeo. Por ejemplo, si los participantes hubiesen hablado entre ellos, quizá se hubiera dado el caso de un cierto consenso respecto a haber juzgado como mentiroso a un determinado declarante. Esto podría haber tenido algún efecto, dado que ese mismo declarante podía volver a aparecer posteriormente. Sin embargo, la precisión no aumentó de forma continua con cada prueba, y a nivel global la precisión obtenida en la cuarta prueba no fue significativamente mayor que la obtenida en la primera. Por lo tanto, es poco probable que los contactos entre los participantes afectaran sustancialmente a los resultados.

En los estudios de Mann y Vrij (2006) y Mann et al. (2007) se emplearon solamente 14 de las 54 declaraciones utilizadas en los trabajos anteriores. Aunque en estas dos investigaciones se presentó a los detectores el mismo número de verdades y mentiras, las limitaciones de las declaraciones a las que hemos hecho referencia son las mismas.

Finalmente, la investigación de Vrij, Akehurst, et al. (2006) también muestra ciertos inconvenientes. En este trabajo el objetivo era examinar la habilidad de profesores, trabajadores sociales, policías y estudiantes para detectar el engaño en niños de 5 y 6 años, adolescentes de 14 y 15 años y adultos. Para ello se crearon tres vídeos distintos. Los participantes realizaron la tarea de

detección distribuidos en grupos, a cada uno de los cuales se les presentaba sólo uno de los tres vídeos (véase Vrij, Akehurst, et al., 2006). Cada vídeo contenía un total de 12 declaraciones, dos verdades y dos mentiras por cada categoría de emisor (niños, adolescentes y adultos). De esta forma, la precisión obtenida al juzgar las declaraciones de los emisores en cada categoría de edad se calculó solamente a partir de cuatro juicios por participante, y los niveles de precisión específicos para las verdades o para las mentiras se basaron solamente en dos juicios por participante.

Otro aspecto a tener en cuenta en este estudio es que en la hoja de respuestas en la que los participantes tenían que reflejar sus juicios se planteaban también otras cuestiones. En concreto, se les preguntaba si el emisor parecía nervioso, si parecía estar haciendo un gran esfuerzo mental, si estaba intentando parecer honesto y si la declaración había proporcionado suficiente información para emitir un juicio. Todas estas cuestiones pudieron haber sido utilizadas por los participantes como base para sus juicios. DePaulo, Lindsay, et al. (2003) señalan que la impresión general de nerviosismo es un indicio realmente útil para detectar el engaño. Además, la investigación ha mostrado que mentir es cognitivamente más complejo que decir la verdad, lo que puede reflejarse en un mayor esfuerzo mental. Finalmente, desde la perspectiva auto-presentacional del engaño (véase DePaulo, 1992) se plantea que un empeño excesivo en mostrarse sincero puede ser un indicio de engaño. En definitiva, las preguntas que cada emisor tenía que contestar tras ver cada declaración pudieron centrar la atención de los participantes sobre algunos indicios válidos de engaño, incrementando así la precisión de sus juicios.

Todas las limitaciones mencionadas de los estudios anteriores, junto con el hecho de que los resultados respecto a la precisión de los profesionales se alejan de lo comúnmente encontrado en el resto de trabajos, invitan a tomar con cierta precaución los datos de tales estudios. De esta forma, si se promedian los datos de las muestras de policías de la Tabla 25 exceptuando las muestras de los seis estudios a los que acabamos de hacer referencia, la precisión media obtenida ($M = 52.03$) se acercaría más al nivel del azar.

Precisión para las declaraciones verdaderas y falsas

Un aspecto importante a tener en cuenta a partir de los datos de los estudios de la Tabla 25 se refiere a la diferencia en la precisión para las verdades y para las mentiras. En general, la investigación que ha trabajado con no profesionales (principalmente estudiantes) ha encontrado unas mayores tasas de precisión al juzgar las declaraciones verdaderas que al juzgar las falsas (Levine et al., 1999). Este dato se podría explicar por una mayor emisión de juicios de verdad que de mentira, es decir, un sesgo hacia la verdad, y así se ha corroborado en distintos trabajos (por ejemplo Bond y DePaulo, 2006; Kalbfleisch, 1985; Vrij, 2000).

Es curioso, sin embargo, observar que esta tendencia parece invertirse en los policías, observándose en la mayoría de ocasiones una precisión superior para las declaraciones falsas que para las verdaderas. Esto sugiere que los policías tienden a hacer más juicios de mentira que de verdad.

En la Tabla 25 se ofrece la precisión para verdades y mentiras por separado en 22 muestras de policías³⁴. Promediando los datos, se obtiene una precisión media para las declaraciones verdaderas ($M = 54.95$) inferior que para las falsas ($M = 60.14$).

Conclusiones respecto a la precisión de los policías

En definitiva, los datos obtenidos a partir de los estudios de la Tabla 25 arrojan las siguientes conclusiones generales: (a) las tasas de precisión de los policías a la hora de distinguir verdades de mentiras se encuentran cerca del nivel del azar; (b) en este sentido, la precisión de los policías es igual de pobre que la que muestran individuos no profesionales; (c) los policías, al contrario que los no profesionales, muestran unas tasas de precisión más elevadas para las declaraciones falsas que para las verdaderas.

Vrij (2004), tras constatar la baja precisión en la detección del engaño, sintetiza en diez puntos las posibles causas de la pobre ejecución de los profesionales:

³⁴ En este caso, no se tienen en cuenta los datos de Vrij, Akehurst, et al. (2006).

a) No existe ningún indicio verbal o no verbal cuya presencia se relacione indefectiblemente con la mentira.

b) Adicionalmente, los indicios que probabilísticamente podrían indicar que alguien está mintiendo generan solamente diferencias muy sutiles entre la conducta de quién miente y la de quien dice la verdad. Tales diferencias no siempre son fáciles de percibir.

c) Según Vrij (2004), es posible que los humanos en ocasiones emitamos nuestros juicios de veracidad basándonos más en reglas de decisión o heurísticos que en un análisis pormenorizado de la conducta y la declaración del emisor, lo que puede llevar a decisiones erróneas. No obstante, no hay una evidencia clara en este sentido cuando se trabaja con policías. Por un lado, distintos estudios (véanse por ejemplo Levine et al., 1999; Stiff, Kim y Ramesh, 1992) plantean que a la hora de juzgar la veracidad es posible que individuos no profesionales basen más sus juicios en reglas de decisión que en un procesamiento pormenorizado. Por otro lado, Masip et al. (2003c) midieron el momento en el que un grupo de policías y otro de estudiantes tomaban sus decisiones sobre la veracidad de un conjunto de declaraciones, y encontraron que los policías tendían más a emitir sus juicios hacia el final de la declaración, mientras que los estudiantes tendían más a hacerlo al principio. Estos resultados podrían indicar que los policías no actuaron bajo la influencia de un heurístico de decisión (en cuyo caso probablemente hubieran emitido antes sus juicios), sino que el hecho de emitir sus juicios hacia el final de la declaración hace pensar que estaban procesando la información que recibían de forma controlada mientras el emisor hacía la declaración, y solamente cuando tenían toda la información (al final de la declaración), tomaban la decisión sobre la veracidad.

d) Vrij (2004) plantea que algunas “reglas conversacionales” podrían dificultar la detección del engaño. Por ejemplo, una regla conversacional establece que lo habitual es mirar a los ojos a la persona con la que se habla. Así pues, un interlocutor que observase otras partes del cuerpo del emisor en busca de indicios de engaño estaría rompiendo esta regla y quizá provocando malestar en el emisor. No hay que olvidar que la investigación empírica ha

mostrado que la detección del engaño a partir de la observación del cuerpo es superior a la obtenida a partir de la observación del rostro (Ekman, 1992; Kalbfleisch, 1985; Masip y Garrido, 2000; Zuckerman et al., 1981).

e) Tal como Vrij indica, y como veremos más adelante con mayor detalle, los indicios en los que los policías basan sus juicios son a menudo aquellos que la investigación ha mostrado que carecen realmente de valor para diferenciar verdades de mentiras (véase también Garrido y Masip, 1999).

f) Otra razón que puede explicar las bajas tasas de precisión es que no se suele prestar atención al contenido verbal de la declaración, que es más revelador que la conducta no verbal (Masip y Garrido, 2000; Zuckerman et al., 1981), especialmente si quien tiene que juzgar la veracidad no está familiarizado con el tema del discurso. Solamente en aquellos casos en los que los policías tienen declaraciones discrepantes de distintas personas sobre un mismo hecho parecen centrarse más en la comparación del contenido. De lo contrario, dirigen su atención principalmente a aspectos no verbales de la conducta.

g) A pesar de que existen importantes diferencias en las formas y estilos en que la gente puede ofrecer una declaración, tanto a nivel verbal como no verbal, a menudo tales diferencias individuales no se tienen en cuenta, y se espera que todo el mundo muestre conductas similares en situaciones similares. No tener en cuenta la idiosincrasia individual en la comunicación puede hacer que se juzgue erróneamente a un inocente como culpable (por ejemplo, si es una persona que muestra claros síntomas de nerviosismo), o a un culpable como inocente (si se trata por ejemplo de una persona muy fría y poco emocional). Ekman (1992) llamó a este fenómeno “riesgo de Brokaw” o “error de idiosincrasia”.

h) Además de la idiosincrasia a nivel personal, hay que tener en cuenta la especificidad de cada situación. Una persona puede comportarse de diferente manera en diferentes situaciones. Pensar que alguien ha de comportarse siempre igual o no tener en cuenta el efecto de las características de la situación puede hacer que se emitan juicios erróneos. Por ejemplo, la conducta

verbal y no verbal de una persona no será la misma cuando hace una primera declaración sobre un hecho que cuando lleva varias horas siendo interrogada.

i) Vrij (2004) también plantea que es posible que existan amplias diferencias individuales entre detectores, siendo algunos profesionales especialmente buenos y otros especialmente malos. Aunque el propio autor admite que es difícil explicar las razones por las que existen estas diferencias individuales entre detectores, estudios como el de Mann et al. (2004) han encontrado una gran variabilidad entre profesionales, que va desde el 30% de precisión en algunos casos al 90% en otros.

j) Finalmente, el autor plantea que las bajas tasas de precisión pueden ser debidas a unas entrevistas y técnicas de interrogatorio policial que no favorecen la detección, sino que más bien pueden introducir sesgos que generan juicios erróneos (véase, por ejemplo, el trabajo de Kassin y Fong, 1999).

LA CONFIANZA EN LA PRECISIÓN DE LOS PROPIOS JUICIOS

Ante estos, quizá sorprendentes, resultados respecto a la precisión de los policías, cabría preguntarse si ellos mismos son conscientes de sus propias limitaciones a la hora de juzgar la veracidad; es decir, ¿en qué medida los policías confían en lo acertado de sus juicios de veracidad? DePaulo, Charlton, Cooper, Lindsay y Muhlenbruck (1997), en un meta-análisis sobre la relación entre confianza y precisión, encontraron que los humanos solemos sobreestimar la precisión de nuestros juicios. Además, hallaron que la relación entre confianza y precisión era prácticamente nula. Este último resultado ha sido replicado recientemente en el meta-análisis de Aamodt y Custer (2006).

En diversos estudios hechos con policías, se han incluido escalas de confianza en los propios juicios. Generalmente, en estos casos, lo que se ha hecho es pedir a los policías que tras cada uno de los juicios de veracidad señalen en una escala tipo Likert la medida en que creen que el juicio que acaban de emitir es correcto. En muchos de estos estudios se ha encontrado que no hay relación entre la confianza y la precisión de los juicios. Además, los policías suelen sobreestimar la precisión de sus juicios (Akehurst et al., 2004;

DePaulo y Pfeiffer, 1986; Ekman y O'Sullivan, 1991; Garrido et al., 2004; Kassin et al., 2005; Mann et al., 2004; Meissner y Kassin, 2002; Vrij y Mann, 2001a, b; Vrij, Mann, Robbins, et al., 2006).

Entre los estudios que han comparado las creencias de los profesionales sobre lo acertado de sus juicios con las creencias de participantes no profesionales, es interesante el de Garrido et al. (2004). En este trabajo, además de encontrar que la precisión de los policías no fue superior a la de los no policías, los autores observaron que tanto policías como no policías consideraban que, en general, los miembros del Cuerpo Nacional de Policía tenían una habilidad para discriminar verdades y mentiras mayor que el ciudadano medio. Pero, además, esta creencia fue más fuerte entre policías que entre no policías. No obstante, no se encontró ninguna correlación significativa entre capacidad estimada y precisión, ni en la muestra de policías ni en la de no policías.

Otros resultados interesantes al respecto son los de Vrij et al. (2007), que examinaron el efecto de tres estilos de entrevista policial sobre la cantidad de indicios obtenidos, la precisión al juzgar las declaraciones y la confianza en los propios juicios de veracidad. Los tres estilos de entrevista considerados fueron:

- a) *Information-Gathering Interview*, que podríamos traducir como entrevista para recoger información, y a la que nos referiremos en adelante como “entrevista facilitadora”. Este tipo de entrevista se caracteriza principalmente por pedir al interrogado que describa los hechos lo más detalladamente posible, utilizando principalmente preguntas abiertas.
- b) *Accusatory Interview*, o “entrevista acusatoria”, en la que los entrevistadores confrontan de forma reiterada al entrevistado con acusaciones directas.
- c) *Behavioral Analysis Interview (BAI)*, que podríamos traducir como “entrevista para el análisis de la conducta”, a la que nos referiremos en adelante como BAI. Esta entrevista consta de dos partes. En un primer momento se empieza planteando al interrogado distintas preguntas abiertas, para pasar después a una segunda fase en la que se formulan

15 preguntas estandarizadas ante las cuales se espera que quienes mientan muestren una conducta diferente que quienes dicen la verdad³⁵.

Los resultados muestran que aunque no hubo diferencias significativas en precisión en función del tipo de entrevista, con la de tipo acusatorio los participantes mostraron una mayor confianza cuando juzgaban una declaración como falsa. Por otra parte, ante declaraciones obtenidas con entrevistas facilitadoras, las tasas observadas de confianza en los juicios fueron las más bajas. De esta forma, el tipo de entrevista que se utiliza para obtener una declaración podría influir en la confianza que los jueces depositan posteriormente en sus juicios. Ante estos resultados, los autores concluyen que los efectos negativos de la baja precisión de los policías a la hora de juzgar la veracidad podrían atenuarse si los propios policías no tuvieran tanta confianza en sus juicios. Plantean que las menores tasas de confianza observadas ante declaraciones obtenidas con entrevistas facilitadoras podrían disuadir a los policías a la hora de emitir juicios de veracidad precipitados, y podrían hacer que dedicasen más tiempo a investigar el caso más a fondo. En definitiva, disminuir la confianza en los juicios puede ser una herramienta que de forma indirecta podría aumentar la precisión.

Finalmente, Kassin et al. (2007) emplearon un cuestionario para formular a 631 policías distintas cuestiones sobre su trabajo policial. Entre otros aspectos, se les pedía que hiciesen una estimación del porcentaje de veces que pensaban que sus juicios de veracidad eran correctos cuando evaluaban la credibilidad de la declaración de un sospechoso. Se halló que los policías estimaban que sus juicios eran correctos un 77% de las veces. Esta estimación es muy superior a las tasas reales a las que antes hemos hecho referencia, y muestra una vez más que los profesionales sobre-estiman su eficacia como detectores de mentiras.

ALGUNAS CONSIDERACIONES FINALES RESPECTO A LA PRECISIÓN

Se han dejado para el final de este apartado una serie de interrogantes que, de forma adicional, podrían plantearse observando los datos de los estudios

³⁵ Más adelante, cuando se trate el tema del entrenamiento que reciben los policías para detectar el engaño, se describirá con más detalle la BAI.

presentados en la Tabla 25. Interrogantes que se refieren tanto a las distintas variables que podrían influir de forma diferencial en la precisión de los policías (tales como las características de los emisores, la modalidad de presentación de las declaraciones o el tipo de entrevista a partir de la que se obtuvieron las declaraciones) como a la posibilidad de la existencia de grupos especialmente hábiles dentro de los profesionales.

¿Existe algún grupo específico de profesionales en el que se haya observado una precisión superior?

En algunas investigaciones se afirma que existen profesionales especialmente hábiles para detectar mentiras. Es el caso, por ejemplo, del trabajo de O'Sullivan y Ekman (2004) al que los autores se refieren como *Wizards Project*, que podría traducirse como "Proyecto de los Magos". O'Sullivan y Ekman utilizan la palabra *wizards* (magos), que también utilizaremos nosotros, para hacer referencia a aquellos individuos que tienen una habilidad especial para juzgar la veracidad de otros, de forma que sus tasas de precisión están muy por encima del resto de las personas.

Para identificar a estos supuestos *wizards* de la detección, los autores plantearon un proceso consistente en tres pruebas diferentes a un total de 12.000 individuos entre los que se incluían terapeutas, personal de las fuerzas de la ley, jueces, abogados, mediadores y artistas. En primer lugar se presentó a los participantes una tarea en la que tenían que juzgar la veracidad de diez declaraciones, cinco verdades y cinco mentiras, en las que el declarante ofrecía su supuesta opinión sobre un tema³⁶. En esta primera prueba se seleccionó solamente a aquellas personas cuya precisión fuera del 90% o superior. A estas personas se les presentaron dos pruebas más de diez declaraciones cada una. En una de ellas se incluían declaraciones en las que los participantes mentían o decían la verdad sobre las emociones que estaban experimentando, y en la otra mentían o decían la verdad sobre si habían robado una cantidad de dinero. El criterio inicial que un individuo debía cumplir para ser calificado como *wizard* era que en estas dos últimas pruebas debía

³⁶ Los dos temas que podían aparecer en las declaraciones eran la prohibición de fumar en lugares públicos y si los delincuentes condenados por asesinato a sangre fría debían ser ejecutados.

obtener una precisión superior al 80%. Siguiendo este procedimiento, se identificaron 14 personas. No obstante, los autores consideraron su criterio de selección demasiado restrictivo, con lo que posteriormente cambiaron las condiciones, de forma que aunque se exigía a los participantes obtener un 90% de precisión en la primera prueba, no era necesario obtener un 80% en las dos restantes; bastaba con llegar a esta puntuación en al menos una de las dos. Con este nuevo sistema, se identificaron 15 *wizards* adicionales.

Una segunda fase en la investigación consistió en volver a visualizar los vídeos con cada uno de los *wizards* identificados. En esta ocasión se les pidió que manifestasen en voz alta las ideas, razonamientos y pensamientos que se les ocurrían respecto a las declaraciones que iban viendo. El objetivo era conocer cómo estas personas elaboran sus juicios de veracidad y en qué se basan para tomar sus decisiones. Siguiendo este método de trabajo, los investigadores llegaron a las siguientes conclusiones: (a) en primer lugar, se encontró una correlación entre la precisión en la detección del engaño y la capacidad para reconocer micro-expresiones faciales de emociones³⁷; (b) quienes alcanzaron una precisión mayor indicaron haber utilizado indicios no verbales, o indicios verbales y no verbales conjuntamente; (c) una estrategia comúnmente utilizada por los *wizards* fue la comparación entre la conducta verbal y no verbal del declarante; (d) en lo demográfico, se observó un predominio de los hombres: sólo 10 de los 29 *wizards* eran mujeres, aunque los propios autores indican que en muchos de los grupos investigados la proporción inicial de hombres era superior a la de mujeres; (e) más curioso es que, según los autores, muchos de los expertos identificados habían tenido algún tipo de experiencia poco usual en la infancia, como ser hijos de alcohólicos o no haber hablado inglés hasta haber superado el periodo de

³⁷ Las micro-expresiones faciales son un tipo de expresiones faciales cuya aparición no se puede controlar y que denotan el verdadero estado emocional de la persona que las muestra. Estas micro-expresiones indican que muchas veces lo que una persona manifiesta de forma controlable, con palabras, no coincide con lo que realmente siente. Quienes mienten tratan de suprimir lo más rápidamente posible la aparición de dichas micro-expresiones, lo que hace que su duración sea muy corta (aproximadamente un cuarto de segundo) y que sean muy difíciles de percibir. Quienes tratan de detectar mentiras podrán beneficiarse de la información de las micro-expresiones (siempre y cuando sean capaces de percibirlas), especialmente cuando se trate de mentiras con alta carga emotiva, para saber qué es lo que realmente está sintiendo la otra persona. El lector interesado en más información sobre micro-expresiones puede consultar los trabajos de Ekman (1985/1992) y Ekman y Friesen (1974).

escolaridad elemental; (f) finalmente, según O'Sullivan y Ekman, todos los *wizards* mostraron haber tenido una motivación muy elevada para detectar el engaño, así como un gran entusiasmo por participar en la investigación.

La idea de O'Sullivan y Ekman (2004) de identificar a los verdaderos especialistas en la detección del engaño, aquellos cuya precisión se sitúa muy por encima del resto de las personas, y posteriormente estudiar las características y estrategias de estos "magos de la detección", sería una forma magnífica de desarrollar protocolos de detección útiles para aquellas personas menos dotadas que, por las características de su trabajo, tienen que decidir si alguien miente o dice la verdad. El problema es que, tal y como Bond y Uysal (2007) han planteado recientemente, quizá tales *wizards* de la detección no existan, y los resultados de O'Sullivan y Ekman (2004) sean fruto del azar. Bondy Uysal (2007) basan su crítica en que la forma considerada por los investigadores para establecer cuantos de los 12.000 participantes podrían ser *wizards* (recordemos, una precisión del 90% en la primera prueba y del 80% en alguna de las otras dos) podría haber determinado la aparición de éstos por mero azar.

Según Bond y Uysal (2007), la consideración de O'Sullivan y Ekman (2004) de que un individuo, por azar, tiene un 50% de probabilidades de juzgar correctamente cada una de las declaraciones (modelo al que estos autores se refieren como "lanzar una moneda al aire") no es una premisa válida para determinar la probabilidad de encontrar *wizards* por azar, ya que cuando se pide a los participantes que juzguen la veracidad de una declaración ellos no toman sus decisiones lanzando una moneda. Como prueba de ello, Bond y Uysal presentan los resultados de distintos trabajos publicados por el grupo de Ekman (véase Bond y Uysal, 2007) en los que muestras de estudiantes habían tenido que evaluar los mismos vídeos que luego emplearían O'Sullivan y Ekman (2004) para identificar a sus *wizards*. La precisión media obtenida por 353, 113 y 464 estudiantes en la detección de mentiras sobre opiniones, haber cometido o no un delito y emociones fue del 56%, del 60% y del 50%, respectivamente. Según estos datos, en los casos de mentiras sobre opiniones y comisión de un delito, los participantes no habrían contestado "lanzando una moneda al aire", pues su precisión está por encima del 50%.

Bond y Uysal (2007) plantean como alternativa el denominado “modelo basado en la investigación”, que permite calcular cuántos individuos de los 12.000 que participaron en el *Wizard Project* podrían haber sido identificados como genios por azar. Para ello se tiene en cuenta tanto la cantidad de participantes como los resultados obtenidos en los trabajos previos. Siguiendo este modelo (véase Bond y Uysal, 2007, para una descripción detallada del mismo), se concluye que, de los 12.000 participantes, sería esperable que alrededor de 70 alcanzasen sólo por azar las puntuaciones necesarias para ser identificados como *wizards*. En definitiva, y tras una serie de críticas metodológicas respecto al desarrollo del *Wizard Project*, tales como haber permitido que fueran los propios participantes los que evaluaran sus propios resultados, o la falta de una definición *a priori* de lo que es un *wizard*, Bond y Uysal (2007) concluyen que el trabajo de O’Sullivan y Ekman (2004) no solamente no ha presentado evidencia consistente de la existencia de estos magos de la detección, sino que tampoco se ha especificado y estandarizado un procedimiento de diagnóstico de tales individuos.

Ante la controversia que se ha suscitado entre O’Sullivan y Ekman por un lado, defendiendo la existencia de una minoría de verdaderos expertos en la detección de mentiras, y Charles Bond y Ahmet Uysal por otro achacando los resultados de los investigadores anteriores al azar, otro investigador, Gary Bond, publicó en el 2007 un trabajo en el que trató de arrojar luz sobre el asunto. En este estudio, un grupo de 112 agentes de la ley (entre los que se incluían policías locales, sheriffs, agentes de la CIA y del FBI, agentes de fronteras y patrulleros) y un grupo de 122 estudiantes realizaron una primera tarea de detección del engaño en la que tenían que evaluar la credibilidad de 32 declaraciones (16 verdades y 16 mentiras). En ocho de ellas el declarante mentía o decía la verdad sobre la comisión de un robo, en otras ocho participaba en una entrevista de trabajo, en otras ocho expresaba los sentimientos hacia otra persona y en las ocho restantes cada entrevistado mentía o decía la verdad sobre si había visto un vídeo (véase Bond, 2007, para una descripción detallada de la elaboración y contenido de los vídeos estímulo).

De los 112 agentes de la ley que participaron en esta primera prueba, 11 obtuvieron una precisión de al menos el 80% o superior (es decir, al menos acertaron 26 de las 32 declaraciones), con lo que fueron seleccionados como posibles expertos y sometidos a una segunda tarea. Ninguno de los estudiantes alcanzó estos niveles de precisión. Para el segundo experimento se pudo contar únicamente con ocho de los 11 profesionales que habían sido seleccionados, a los que se añadió un grupo control de otros ocho estudiantes. En este caso se utilizaron 10 nuevas declaraciones (5 verdades y 5 mentiras) sobre los mismos temas del experimento anterior. De los ocho supuestos expertos, solamente dos mostraron una precisión superior al 80% (en concreto, la precisión global de ambos fue del 90%).

Al final del estudio, el autor se pregunta si los resultados podrían explicarse por azar. Sin contestar de forma tajante a esta pregunta, lo que Gary Bond ofrece es un cálculo de la probabilidad de que un participante acertase por azar 39 del total de los 42 ítems que se emplearon en los dos experimentos. Según el autor, dicha probabilidad, teniendo en cuenta el total de participantes y las tasas medias obtenidas, sería de .000002 (véase Bond, 2007). A renglón seguido, el propio investigador presenta las limitaciones principales de este trabajo, entre las que destaca el hecho de que solamente 10 personas generaron las 32 declaraciones del primer experimento y las 10 del segundo, con lo que, necesariamente, una misma persona aparecía en más de una ocasión, y posiblemente en los dos experimentos. Además, el propio autor plantea la necesidad de que, para que un participante fuera calificado como experto, sería necesario que pasase por más pruebas. Este es un detalle importante, ya que si de los ocho expertos que pasaron la segunda prueba solamente dos obtuvieron puntuaciones superiores al 80%, ¿qué podría haber pasado en una tercera prueba?; ¿cómo decidir cuantas pruebas son necesarias para poder concluir con seguridad de que nos hallamos ante un experto?

Hasta el momento, pues, la pregunta que encabezaba este epígrafe, referida a la existencia entre los profesionales del ámbito legal de algún grupo en el que, de forma sistemática, se haya observado una precisión superior, no tiene una respuesta clara. Quizá sería necesario, en primer lugar, que los distintos

investigadores se pusieran de acuerdo y desarrollasen un protocolo estandarizado para identificar verdaderos expertos. Una vez dado este primer paso, habría que esperar a que hubiese evidencia suficiente para establecer una conclusión con cierta solidez.

¿Son los profesionales igualmente poco eficaces cuando juzgan la veracidad de los adultos y de los niños?

Aunque lo más común es que la precisión de los policías se estudie con grupos de adultos como emisores, en algunos casos se han utilizado niños. Los trabajos de Leach et al. (2004) y de Vrij, Akehurst, et al. (2006) son un ejemplo de ello.

Leach et al. (2004) presentan tres experimentos en los que policías, agentes de aduana y estudiantes tenían que juzgar la credibilidad de un grupo de declaraciones hechas por niños de entre 3 y 11 años. Los investigadores emplearon un paradigma de resistencia a la tentación en el cual cada niño participaba en un proceso de interacción con una experimentadora. En un momento determinado, ésta abandonaba la habitación e indicaba al niño que no debía mirar un juguete determinado que había en la sala. Algunos de los niños obedecían la orden de la experimentadora, y otros no. Cuando ésta regresaba a la habitación, hacía tres preguntas a los niños: (a) “¿mientras no estaba, moviste tu cabeza hacia un lado?”, (b) “¿te moviste alrededor de la silla?”, y (c) “¿curioseaste para ver cuál era el juguete?” Algunos niños mintieron y otros dijeron la verdad sobre si habían o no mirado el juguete. Las declaraciones de los niños se grabaron en vídeo y se presentaron a los detectores. En esta condición, la precisión de los tres grupos (policías, agentes de aduana y estudiantes) fue, respectivamente, del 44%, del 49% y del 51%. Las diferencias entre agentes de aduana y estudiantes no fueron significativas. Ambas puntuaciones no se diferenciaron significativamente del 50% esperado por azar. La precisión de los policías fue significativamente inferior a la de los otros dos grupos, y además estuvo por debajo de lo esperado por azar.

En el segundo estudio, Leach et al. (2004) utilizaron un paradigma similar al anterior, pero se introdujo una modificación. Previamente a que los niños

contestasen a las preguntas de la experimentadora, se discutió con ellos la moralidad de mentir o decir la verdad. Adicionalmente, con el objetivo de aumentar el tiempo de duración de cada declaración, se incluyó una muestra de la conducta de la interacción del niño con la experimentadora en la que ésta hacía preguntas al niño no relacionadas con la posible transgresión, de forma que cada declaración constaba de dos partes: (a) muestra de la conducta previa del niño, y (b) respuestas a las preguntas sobre la posible transgresión. Con esta modificación, las tasas de precisión fueron del 64%, del 62% y del 63% para policías, agentes de aduana y estudiantes, respectivamente. Todas las puntuaciones estuvieron por encima del nivel del azar, y las diferencias entre los tres grupos no fueron significativas. Los resultados de este segundo estudio pueden reflejar la utilidad de exponer a los detectores previamente a muestras de la conducta de los emisores en las que éstos son sinceros. Alternativamente, es posible que tras la discusión sobre la moralidad de mentir o decir la verdad los niños emitieran mentiras más transparentes.

En el tercer estudio, el contexto experimental y el formato de las declaraciones fue similar al del estudio 2, pero en lugar de discutir con los niños la moralidad de mentir o decir la verdad se les hizo prometer, antes de contestar a las preguntas de la experimentadora, que dirían la verdad. Las tasas de precisión fueron del 59%, del 57% y del 61% para policías, agentes de aduana y estudiantes, respectivamente. De nuevo, las diferencias entre los tres grupos no fueron significativas y todos tuvieron una precisión superior al azar.

Vrij, Akehurst, et al. (2006), por su parte, realizan un interesante estudio en el que comparan las tasas de precisión de policías, trabajadores sociales, profesores y estudiantes al juzgar las declaraciones de niños (de 5 y 6 años), adolescentes (de 14 y 15 años) y adultos. Los investigadores encontraron que no había diferencias significativas en función del grupo de pertenencia de los observadores. En todos los casos, la precisión estuvo en torno al 60%. Respecto a los grupos de emisores, solamente se encontró una precisión superior cuando se juzgaron declaraciones verdaderas ofrecidas por adolescentes (67%). En los demás casos, como se puede observar en la Tabla 25, la precisión se situó en torno a un 60%.

En definitiva, tanto el trabajo de Leach et al. (2004) como el de Vrij, Akehurst, et al. (2006) parecen indicar que al juzgar la veracidad de declaraciones hechas por niños la precisión de profesionales y no profesionales es similar a la obtenida al juzgar declaraciones de adultos.

¿Influye el modo de presentación de las declaraciones en la precisión de los policías?

Aunque en la mayoría de los casos se ha utilizado un formato audiovisual, algunos estudios han tratado de establecer si había diferencias en precisión en función de si los policías solamente veían o solamente escuchaban las declaraciones que tenían que juzgar. Kassin et al. (2005) realizaron un estudio en el que establecieron dos condiciones experimentales: vídeo y audio. En la condición vídeo (imagen sin sonido), un grupo de 28 policías y otro de 29 estudiantes tuvieron que juzgar la veracidad de 10 declaraciones (cinco verdaderas y cinco falsas) sobre distintos delitos. La precisión de los estudiantes fue del 53.4% y la de los policías del 42.1%, inferior a la de los estudiantes, y al 50% esperado por azar. En la condición audio, 29 estudiantes y 32 policías juzgaron la veracidad de las mismas 10 declaraciones, pero esta vez, en lugar del sonido, se eliminó la imagen. Las tasas de precisión fueron del 64.1% para los estudiantes y del 54.5% para los policías. De nuevo, la precisión de los estudiantes es superior a la de los policías, que se sitúa en torno al nivel del azar. En este estudio se encontraron efectos principales significativos tanto de la muestra como del modo de presentación. En concreto, los estudiantes fueron más precisos que los policías, y la precisión (para ambos grupos) fue mayor en formato audio que en formato vídeo.

En otro estudio posterior, Mann et al. (2007) compararon la precisión de los policías en tres condiciones diferentes: vídeo, audio, y vídeo más audio. En la condición de vídeo, 37 policías visualizaron 14 *clips* (siete verdades y siete mentiras), en cada uno de los cuales aparecía un individuo haciendo una declaración a la que se le había quitado el sonido. Obtuvieron una precisión global del 43% (34% para las verdades y 53% para las mentiras). En la condición audio, 31 policías fueron expuestos a los mismos 14 *clips*, pero esta vez se había eliminado la imagen, con lo que solamente podían oír la

declaración. En esta modalidad, la precisión global fue superior, del 66% (63% para las verdades y 69% para las mentiras). Los autores encontraron que las diferencias entre estas dos condiciones fueron significativas, es decir, los policías tenían una mayor precisión al juzgar la veracidad cuando las declaraciones se presentaban en formato audio. En la condición audio más vídeo, se presentaron las mismas 14 declaraciones a un grupo de 35 policías, pero en este caso podían ver al declarante y también escuchar lo que decía. Bajo esta condición, la precisión global de los juicios fue del 65% (60% para las verdades y 70% para las mentiras). Estos resultados no se diferenciaron de los encontrados en la modalidad audio, pero sí fueron mejores que los de la modalidad vídeo.

En definitiva, estos dos estudios parecen indicar que la precisión de los juicios de veracidad de los policías mejora cuando, además de ver la declaración, también pueden escuchar lo que dice el emisor. Los resultados respecto a la modalidad de presentación encontrados para los policías van en la misma línea que los resultados de los estudios hechos con no profesionales, y la cuantía de las tasas de precisión es similar. En el meta-análisis de Kalbfleisch (1985), se observó una precisión del 51% para la modalidad vídeo, 58% para la modalidad audio y 57% para la modalidad audiovisual. En el recientemente meta-análisis de Bond y DePaulo (2006), los resultados son similares: la precisión para la modalidad vídeo fue del 52%, para la modalidad audio del 62% y para la condición audiovisual del 56%.

Conclusiones respecto a los tres interrogantes anteriores

De los trabajos revisados en los epígrafes anteriores se pueden extraer las siguientes conclusiones generales: (a) no está claro que exista entre los profesionales un grupo de individuos (a los que nos hemos referido como *wizards*) especialmente habilidosos a la hora de detectar el engaño (es decir, cuyas tasas de precisión estén muy por encima de las de otros profesionales); (b) las tasas de precisión de los policías son similares a las de los no profesionales, tanto cuando juzgan la veracidad de declaraciones hechas por personas adultas, como cuando juzgan la veracidad de declaraciones emitidas por jóvenes o niños; y (c) en general, parece que la precisión de los policías,

como sucede con la de observadores legos, es mayor cuando las declaraciones se presentan en formato audio o audiovisual que cuando se presentan en formato visual.

EL INVESTIGATOR BIAS EFFECT

Tratar de dar explicación a las dos conclusiones del apartado anterior que más relevancia pueden tener para el desarrollo de la labor policial (la pobre precisión de los policías y su tendencia a juzgar las declaraciones como falsas), ha sido el objetivo de muchos investigadores. Dos factores que pueden tener una influencia decisiva sobre los juicios de veracidad de los policías son la experiencia en el trabajo policial y el entrenamiento que estos profesionales reciben para detectar mentiras. En este apartado expondremos distintos trabajos, centrándonos en la investigación de Meissner y Kassin (2002) sobre el efecto de estas dos variables, ya que ha supuesto el punto de partida para los estudios realizados en esta Tesis Doctoral.

LA INVESTIGACIÓN DE MEISSNER Y KASSIN

Gran parte de la investigación de S. M. Kassin en Estados Unidos se ha centrado en estudiar cómo, en determinadas ocasiones, las tácticas que los policías suelen utilizar en sus interrogatorios pueden hacer que algunas personas inocentes confiesen haber cometido aquellos delitos de los que se les acusa. Es lo que se conoce como falsas confesiones. En esta línea están algunas de las más recientes publicaciones de este autor (por ejemplo, Kassin y College, 2005; Kassin et al., 2007; Kassin et al., 2005), y otros trabajos anteriores en los que Kassin advierte también del peligro de los sesgos que los distintos profesionales pueden llevar consigo a la sala de interrogatorios o a los tribunales (por ejemplo, Kassin, 2002; Kassin, Goldstein y Savitsky, 2002). En esta línea de trabajo, Meissner y Kassin (2002) llevan a cabo un pequeño meta-análisis en el cual se muestra: (a) que las tasas de precisión de los policías al juzgar la veracidad son muy bajas; y (b) que tanto el hecho de ser policía como haber recibido entrenamiento para detectar mentiras resulta en una tendencia a juzgar las declaraciones como falsas. A este segundo hallazgo los autores lo denominan *investigator bias effect* (efecto del sesgo del

investigador). Además, Meissner y Kassin (2002) deciden realizar un estudio empírico sobre el *investigator bias effect*.

Los autores tomaron los datos de un estudio anterior (Kassin y Fong, 1999), en el cual 40 estudiantes habían sido aleatoriamente asignados a un grupo de entrenamiento o a un grupo sin entrenamiento. Ambos grupos debían juzgar la veracidad de 16 declaraciones (8 verdades y 8 mentiras) grabadas en vídeo (véase Kassin y Fong, 1999, para una descripción detallada de las declaraciones). La manipulación experimental consistió en que, a diferencia del grupo control de no entrenamiento, los 20 participantes del grupo de entrenamiento fueron entrenados con la *Técnica Reid* para la detección de mentiras antes de realizar la tarea. Esta técnica es una de la más utilizadas en Estados Unidos para entrenar a los policías en la detección del engaño.

Tras analizar los datos recogidos de la tarea de detección, Kassin y Fong (1999) encontraron que aunque el entrenamiento no mejoraba las tasas de precisión (46% y 56% para el grupo de entrenamiento y el grupo sin entrenamiento, respectivamente), los participantes entrenados mostraron una confianza en sus juicios significativamente mayor que los no entrenados (véase Kassin y Fong, 1999).

Meissner y Kassin (2002) reanalizaron estos datos desde el marco de la Teoría de la Detección de Señales, con el fin de estimar la capacidad de discriminación (A'), y el sesgo de respuesta (B'')³⁸. Los resultados de estos nuevos análisis mostraron que quienes recibieron entrenamiento ($A' = .52$) discriminaron verdades y mentiras peor que quienes no fueron entrenados ($A' = .63$). El tamaño del efecto encontrado ($d = -.45$) muestra que la diferencia a favor del grupo no entrenado fue notable. Respecto a la existencia de un sesgo de respuesta, los individuos entrenados mostraron un leve sesgo hacia la mentira ($B'' = -.18$), mientras que quienes no fueron entrenados mostraron un leve sesgo hacia la verdad ($B'' = .14$). En general, estos resultados ofrecen apoyo a la hipótesis de que el entrenamiento no mejora la precisión, sino que

³⁸ El rango de A' abarca de 0 a 1. Cuanto más se acerque A' a 1, mejor será la discriminación. B'' abarca de -1 a 1. Los valores negativos indican un sesgo hacia la mentira, los positivos un sesgo hacia la verdad.

simplemente aumenta la probabilidad de que una declaración sea juzgada como falsa.

Meissner y Kassin (2002) habían hipotetizado también que la experiencia como profesional, lejos de mejorar la precisión, generaría un sesgo hacia la mentira. Para contrastar esta hipótesis compararon los datos de los dos grupos de estudiantes de Kassin y Fong (1999) con los obtenidos con 44 policías. Su experiencia media era de 13.7 años de trabajo policial, y el 68% de ellos había recibido entrenamiento profesional sobre entrevistas, interrogatorios o detección de la mentira. Se les presentó el mismo conjunto de declaraciones que habían visto los dos grupos de estudiantes, e igualmente se les pidió que juzgasen la veracidad de cada una de ellas.

Al comparar los resultados de los policías con los de los estudiantes se observó que la precisión de los policías ($M = 50\%$) no se diferenciaba de la de los estudiantes entrenados ($M = 46\%$) ni de la de los estudiantes no entrenados ($M = 56\%$). Sin embargo, tanto los estudiantes entrenados como los no entrenados mostraron niveles de confianza inferiores a los de los policías. Los análisis realizados desde la Teoría de la Detección de Señales indicaron que los tres grupos no diferían en su nivel de discriminación. Sin embargo, tanto los policías ($B'' = -.46$) como los estudiantes entrenados ($B'' = -.18$) mostraron un claro sesgo a juzgar las declaraciones como falsas, mientras que los estudiantes no entrenados mostraron una tendencia en sentido opuesto ($B'' = .14$).

Finalmente, Meissner y Kassin (2002) examinaron la existencia de correlaciones entre entrenamiento y experiencia de los policías, y la precisión y la confianza en sus juicios. En primer lugar, se observó que los años de experiencia en el Cuerpo de Policías no se relacionaban ni con la precisión en los juicios ni con la confianza. Se encontró, sin embargo, que una mayor experiencia como policía se relacionaba con un incremento del sesgo hacia la mentira. Respecto al entrenamiento, los resultados fueron similares, ya que no correlacionaba significativamente ni con la precisión ni con la confianza en los juicios, pero sí se encontró una correlación marginalmente significativa entre la cantidad de entrenamiento y el sesgo hacia los juicios de mentira.

A partir de estos resultados, Meissner y Kassin (2002) consideran probada la existencia de un sesgo del investigador, que se manifiesta en la tendencia de los policías a juzgar las declaraciones como falsas. Este sesgo, según los autores, se ve facilitado por la experiencia como profesional y por el entrenamiento recibido.

Al inicio de esta Tesis Doctoral, se planteaba como objetivo la profundización en el estudio del efecto del sesgo del investigador. Con el fin de establecer un orden en la exposición del análisis de este sesgo, se abordarán por separado las dos variables que están en su base: la experiencia policial y el entrenamiento. Comenzaremos, en primer lugar, con la experiencia policial, presentando los hallazgos de los estudios que han servido de base a las hipótesis que se postulan desde esta Tesis sobre el efecto de dicha variable. Se presentarán, a continuación, los resultados encontrados en la investigación que hemos realizado. En segundo lugar, abordaremos el tema del entrenamiento, examinando la investigación que ha estudiado su efecto. Nos referiremos tanto al entrenamiento que pueden recibir los policías durante su formación, como al recibido por otras muestras en estudios experimentales. Presentaremos las hipótesis que planteamos respecto al entrenamiento, y describiremos la investigación que hemos realizado al respecto.

El efecto de la experiencia profesional sobre los juicios de veracidad de los policías

6

¿A través de qué mecanismos puede influir la experiencia profesional en los juicios de veracidad de los policías? ¿Cómo actúa esta variable para no aumentar la precisión de los juicios de veracidad y generar, junto con el entrenamiento, un sesgo hacia la mentira? Estas son las preguntas que han servido de base para la investigación de esta Tesis sobre el efecto de la experiencia policial.

Una posible explicación del sesgo hacia la mentira en los juicios de los policías puede ser el hecho de que éstos, ante una tarea de evaluación de la veracidad, traten deliberadamente de buscar indicios de mentira, es decir, traten de confirmar una expectativa previa de que el individuo que hace la declaración está mintiendo. Esta posibilidad ha encontrado apoyo, por ejemplo, en los resultados del estudio de Garrido et al. (2004; véanse también Garrido et al., 2000, y Masip, Garrido, Herrero, Antón, et al., 2006), en que se comparó la ejecución de un grupo de policías y otro de estudiantes en una tarea de evaluación de la veracidad. Tras analizar los datos, los autores encontraron que aunque la precisión global media (53%) no difería significativamente de lo esperado por azar, los policías tendían a juzgar la mayor parte de las declaraciones como falsas con independencia de si eran realmente falsas o no, lo que hizo que fueran mucho menos precisos que los no-policías al detectar las declaraciones verdaderas. En contraste, los estudiantes fueron más sensibles a la verdadera naturaleza de las declaraciones.

Cuando los investigadores pidieron a los participantes que indicasen las claves a las que habían atendido para hacer sus juicios, encontraron que, pese a haber visto las mismas declaraciones, los policías y los no-policías parecían haber percibido cosas opuestas en ellas. Los primeros dijeron haber detectado conductas que asociaban a juicios de engaño (conducta nerviosa, conductas de inhibición, dudas o inseguridades al narrar, implausibilidad, evitación del

contacto ocular y contradicciones verbales); los estudiantes mencionaron conductas opuestas, que además asociaron a juicios de verdad (no-nerviosismo, conducta relajada, confianza de la testigo al declarar, plausibilidad, mantenimiento del contacto ocular y consistencia verbal). Ante estos resultados, Masip, Garrido, Herrero, Antón, et al. (2006) concluyen que los policías pueden tener una preconcepción de que el testigo miente, de forma que atienden selectivamente a claves asociadas a juicios de mentira, ignoran aquellas que se asocian a juicios de verdad e interpretan conductas ambiguas como indicadores de mentira. Es decir, muestran un sesgo confirmatorio.

¿Qué mecanismo podría explicar la existencia de un sesgo confirmatorio en los juicios de los policías, tal como Masip, Garrido, Herrero, Antón, et al. (2006) proponen? Desde esta Tesis, se defiende que la formación de este sesgo se produce de forma progresiva. Para explicar esta idea es necesario tomar como punto de partida el trabajo de Levine y McCornack (1991) sobre la sospecha. Estos autores diferencian entre Sospecha Comunicacional Generalizada (*Generalized Communicative Suspicion* o GCS) y Sospecha Situacional o Estado. La primera (GCS) se entendería como una “predisposición a creer que los mensajes generados por los demás son mentiras” (Levine y McCornack, 1991, p. 328). La GCS se concibe así como un constructo cognitivo relativamente estable a través del tiempo y de las situaciones, y podría compararse en cierto modo con un rasgo de personalidad. Por el contrario, la sospecha estado o situacional se genera ante determinadas claves contextuales, y se entiende como “la creencia de que una comunicación dentro de un contexto determinado y en un momento determinado puede ser una mentira” (Levine y McCornack, 1991, p. 328). A diferencia de la GCS, la sospecha estado es transitoria y tiene su origen en uno o varios factores de la situación. A continuación se presentarán algunos estudios que hacen referencia a los efectos de la sospecha sobre los juicios de veracidad.

EVIDENCIA EMPÍRICA RESPECTO A LA SOSPECHA

Toris y DePaulo (1984) examinaron cómo la sospecha estado afecta a la precisión en la detección del engaño y a la percepción de los otros como mentirosos o sinceros. Cuarenta individuos tenían que entrevistar durante un

periodo aproximado de 5 minutos a distintas personas en referencia a su personalidad y debían tratar de formarse una imagen lo más correcta posible sobre el grado en que el entrevistado era introvertido o extrovertido. La sospecha se manipuló diciendo a la mitad de los entrevistadores que algunas de las personas que iban a entrevistar tratarían de mostrar una falsa imagen de sí mismas. Después de cada entrevista, se pedía al entrevistador que indicase como creía que era la personalidad del entrevistado. Adicionalmente se le presentaban una serie de escalas tipo *Likert* de siete puntos en las que se le indicaba que valorase al entrevistado en referencia a distintos aspectos, entre los que se incluía su sinceridad durante la entrevista 1 (*completamente sincero*), y 7 (*completamente insincero*). Finalmente, se incluyó también otra pregunta sobre la seguridad en el propio juicio.

Tras analizar los datos recogidos, se encontró que no hubo diferencias significativas en precisión en la detección del engaño entre entrevistadores a los que se les había inducido sospecha y aquellos a los que no. Sin embargo, los entrevistadores en la condición de alta sospecha estado mostraron una tendencia marginalmente significativa ($p = .07$) a percibir a las personas entrevistadas como más mentirosas, con independencia de si éstas mentían o decían la verdad. Además, fueron menos confiados en sus juicios de veracidad. Las autoras concluyen que la sospecha estado no incrementa la precisión de los juicios, sino que hace que la persona se vuelva más desconfiada, perciba a los demás como menos sinceros, y reduzca la seguridad en sus propios juicios.

Levine y McCornack (1991) diferenciaron entre sospecha estado y GCS, y en un primer estudio elaboraron una escala para medir este último constructo. El objetivo, en este caso, era valorar la calidad de los ítems, la fiabilidad de la escala y establecer la unidimensionalidad y validez de constructo. Un total de 141 estudiantes completaron 14 ítems que inicialmente componían la escala de GCS (ver Anexo E) junto con otros tres cuestionarios³⁹. Se tomaron tres criterios a la hora de decidir qué ítems se conservarían: (a) que el peso factorial de cada ítem fuese superior a .40, (b) que cada ítem contribuyera positivamente a la fiabilidad de escala y (c) que los ítems fueran consistentes

³⁹ Estos tres cuestionarios medían agresividad, aceptación de la mentira y aprensión comunicativa (véase Levine y McCornack, 1991, para una explicación más detallada).

con un modelo factorial unidimensional. Teniendo en cuenta estos requisitos, se retuvieron 6 de los 14 ítems. La fiabilidad de la escala (Alfa =.71) fue, según los autores, aceptable.

En un segundo estudio se planteó como objetivo examinar si la escala de GCS correlacionaba positiva o negativamente con otras escalas ya validadas que medían constructos similares. Adicionalmente, este estudio sirvió para replicar los análisis respecto a la estructura factorial de la escala. En este caso, 159 estudiantes completaron la escala junto con veinticinco ítems de la escala de confianza interpersonal de Rotter (1967), cuatro de la subescala de cinismo de la escala de filosofía de la naturaleza humana y tres de la subescala de negativismo de la escala de maquiavelismo de Christie y Geis (1970), 14 de la subescala de veracidad de Wrightman (1964) y ocho ítems de la escala de confianza específica de Larzelere y Huston (1980). Siguiendo los criterios de conservación de ítems establecidos en el primer estudio, de los 14 ítems de la escala de GCS, solamente tres fueron desechados. La fiabilidad (Alfa = .82) fue superior a la del primer estudio. Se observó que la escala de GCS correlacionaba significativa y positivamente con las subescalas de cinismo y la de negativismo. Además, correlacionaba significativa y negativamente con las escalas de confianza interpersonal, veracidad y confianza específica. Sin embargo, las correlaciones no eran tan altas como para considerar estos constructos isomórficos con la GCS. Estos resultados, junto con los datos del primer estudio, confirman según Levine y McCornack (1991) que por un lado la escala de GCS mide un constructo específico diferente de otros relacionados, y por otro, que las cualidades psicométricas de la escala hacen de ella un instrumento fiable y válido.

En un tercer estudio se trató de comprobar la utilidad predictiva de la escala de GCS. En concreto, se esperaba que quienes obtuvieran mayores puntuaciones en dicha escala, en una tarea posterior de evaluación de la veracidad mostrarán un sesgo hacia la mentira, es decir, emitirían una mayor cantidad de juicios de mentira. Además, se esperaba que la relación entre GCS y la cantidad de juicios de mentira se mantuviera constante al controlar el efecto de la sospecha estado.

En este estudio participaron 107 parejas. Uno de los miembros de cada pareja adoptó el rol de juez y el otro el de emisor. La tarea de los jueces consistía en juzgar la veracidad de las respuestas del emisor a 12 cuestiones sobre su personalidad (seis de las respuestas eran verdaderas y las otras seis falsas). Los emisores se asignaron de forma aleatoria a la condición de sospecha estado baja (a los jueces no se les dio ninguna información sobre la posibilidad de que quienes hacían las declaraciones pudieran mentir), moderada (el experimentador informó a los jueces de que existía la posibilidad de que en algunas de las declaraciones que iban a ver los declarantes no fueran del todo sinceros), o alta (se informó a los jueces de que algunas de las declaraciones eran completamente falsas, y se les dijo que su tarea era determinar cuales eran). Además, cada juez cumplimentó la escala de GCS. Las puntuaciones obtenidas se dividieron en tres grupos: baja, media y alta GCS.

Los resultados de este estudio indican que, globalmente, a mayor GCS más juicios de mentira se hicieron, aunque las diferencias entre los niveles intermedio y alto de GCS no fueran significativas. Así pues, Levine y McCornack (1991) consideran que estos datos proporcionan apoyo a la utilidad predictiva de la escala de GCS. Sin embargo, este efecto dependía del nivel de sospecha estado. Cuando ésta era baja, quienes tenían niveles moderados o altos de GCS hicieron más juicios de mentira que quienes tenían niveles bajos de GCS. Cuando la sospecha estado era intermedia, quienes tenían niveles altos de GCS hicieron más juicios de mentira que quienes los tenían intermedios, y éstos más que quienes los tenían bajos. Finalmente, cuando la sospecha estado era baja, quienes tenían niveles de GCS moderados hicieron más juicios de mentira que quienes los tenían bajos o altos.

Una debilidad de este estudio la constituye el pequeño tamaño de la muestra. Al cruzar los tres niveles de sospecha estado con los tres niveles de GCS se obtiene un diseño con nueve celdas. La cantidad de participantes con que se contó en cada celda fue muy pequeña, menos de 10 en tres celdas, 10 en dos, y entre 11 y 15 en cuatro. Quizá con más participantes podría establecerse una clasificación de las puntuaciones en GCS en más de tres niveles y si, tal como plantean Levine y McCornack (1991), el efecto de la GCS

puede que no sea muy grande, se observarían mejor las diferencias cuanto más extremas fuesen las puntuaciones.

Posteriormente, McCornack y Levine (1990) desarrollaron una investigación en la que trataron de comprobar si distintos niveles de sospecha estado se relacionaban de forma diferencial con la precisión de los juicios de veracidad, y si la relación sospecha estado – precisión se veía moderada por los niveles de GCS (altos o bajos) propios de cada individuo. En este estudio participaron 107 parejas. Uno de los miembros de cada pareja adoptaba el rol de juez de las declaraciones del otro⁴⁰, que se le presentaban en formato audiovisual. La GCS de tales jueces se midió con un cuestionario diseñado a tal efecto, y se dicotomizó la muestra diferenciando entre personas con alta GCS y personas con baja GCS. Además se establecieron tres niveles de sospecha estado: bajo, moderado, y alto. Esta variable se manipuló como lo hicieron McCornack y Levine (1991) en el trabajo antes revisado.

Los resultados de este estudio mostraron que, en general, la mayor precisión se obtuvo con niveles intermedios de sospecha estado. Además, considerando el efecto de la GCS, la precisión de los participantes del grupo de baja sospecha estado fue similar entre quienes habían obtenido altas o bajas puntuaciones en la escala de GCS. En los grupos de sospecha estado moderada y alta los que alcanzaron una mayor precisión fueron quienes habían mostrado una mayor GCS.

Stiff et al. (1992) se propusieron investigar la influencia de la sospecha estado y el sesgo hacia la verdad sobre los juicios de veracidad en una tarea de detección del engaño. Para ello contaron con 80 parejas de amigos íntimos. Uno de los miembros de cada pareja adoptaba el rol de juez, y su tarea era evaluar, a través de una entrevista, la veracidad de las declaraciones del otro miembro, el cual tenía que responder a distintas preguntas sobre las reacciones emocionales que había experimentado al ver dos *clips* de vídeo.

⁴⁰ De cada declarante se grabaron en vídeo 12 segmentos, cada uno de los cuales se correspondía con la respuesta a uno de 12 ítems aleatoriamente extraídos de una escala de maquiavelismo. En seis segmentos el declarante decía la verdad y en los otros seis su respuesta era falsa.

Se establecieron dos niveles de sospecha estado. A los jueces asignados al grupo de alta sospecha estado, se les dijo (previamente a la tarea de detección) que a algunos de los emisores se les había pedido que contestaran sinceramente durante la entrevista, pero que a otros se les había pedido que mintiesen. A los jueces asignados al grupo de baja sospecha estado no se les dio ninguna indicación respecto a la potencial veracidad de la persona a la que iban a entrevistar.

Stiff et al. (1992) esperaban que los jueces asignados al grupo de baja sospecha estado mostraran un mayor número de juicios de verdad, debido a la presencia de un sesgo de veracidad resultante de la relación de amistad entre los dos miembros de la pareja. Por otro lado, se esperaba que el sesgo de veracidad de los participantes del grupo de alta sospecha estado fuera menor debido al efecto de la sospecha inducida. Tras analizar los datos, se encontró que, efectivamente, el grupo de baja sospecha estado efectuó una mayor cantidad de juicios de verdad que el grupo de alta sospecha estado. De hecho, los integrantes del grupo de alta sospecha estado no mostraron un sesgo hacia la verdad. No obstante, la cantidad de juicios de verdad no tuvo una relación significativa con la precisión de los juicios en ninguno de los dos grupos. Así pues, al igual que en el estudio de McCornack y Levine (1990), la sospecha estado aumentó la cantidad de juicios de mentira eliminando el sesgo hacia la verdad.

En el estudio de Millar y Millar (1997), 107 participantes tuvieron que juzgar si una serie de personas mentían o decían la verdad al describir una serie de fotografías. En la condición de alta sospecha estado se dijo a los jueces que algunos emisores podían introducir información falsa al hacer las descripciones. En la condición de baja sospecha estado no se dio esta información.

Tras analizar los datos, se observó que los participantes del grupo de alta sospecha estado hicieron significativamente más juicios de mentira que los del grupo de baja sospecha estado, que mostraron un sesgo hacia la verdad. En consecuencia, la precisión para las declaraciones falsas fue superior en la condición de alta sospecha estado que en la condición de baja sospecha, pero

para las declaraciones verdaderas el patrón fue el inverso. No obstante, la precisión global de los dos grupos de jueces fue similar. De nuevo, la sospecha aumentó la cantidad de juicios de mentira, pero no mejoró la precisión.

Un aspecto importante de este estudio es que los autores quisieron comprobar si cuando los detectores disponían de menos recursos cognitivos para elaborar sus juicios se veían más influenciados por el sesgo que se les hubiera inducido, ya fuera de verdad (en el caso del grupo de baja sospecha estado), o de mentira (en el caso del grupo de alta sospecha estado). Para ello, los investigadores establecieron, dentro de cada uno de estos grupos, una manipulación adicional, de tal manera que en ambos grupos los jueces observaron la mitad de las declaraciones mientras hacían una serie de problemas aritméticos, y la otra mitad mientras disponían del cien por cien de sus recursos cognitivos.

Los resultados mostraron que, en condiciones de baja sospecha estado, el sesgo de veracidad era más pronunciado cuando los jueces tenían sus recursos cognitivos limitados. El sesgo de veracidad pudo actuar como una regla de decisión heurística ante la falta de recursos cognitivos. Curiosamente, en el grupo de alta sospecha estado, no hubo diferencias en función de los recursos cognitivos disponibles. Es decir, tener la posibilidad de procesar los datos más profunda y detalladamente cuando se ha inducido la sospecha no hace que se reduzca el sesgo de mentira generado por dicha sospecha⁴¹. Es posible, sin embargo, que la sospecha pudiese redirigir dichos recursos hacia un procesamiento sesgado, tratando de confirmar que el emisor miente, de forma que, tal como indican Masip, Garrido, Herrero, Antón, et al. (2006), se intente identificar conductas que se asocian a juicios de engaño. A partir de este razonamiento se podría sugerir que la sospecha estado no hace que disminuya el procesamiento de los datos entrantes, sino que lo “contamina”.

⁴¹ Podría pensarse que el hecho de disponer de más recursos cognitivos, posibilitaría al detector un análisis más profundo y detallado de la conducta del emisor, que tendría como resultado unos juicios más basados en el escrutinio de dicha conducta que en la sospecha inducida, con lo que en estas condiciones se esperaría que se redujese el número de juicios de mentira.

Bond y Lee (2005b) trataron de comprobar la fiabilidad y validez de la escala de GCS de Levine y McCornack (1991). Para ello realizaron dos estudios con reclusos. En el primero, un total de 80 reclusos de entre 18 y 43 años completaron cuatro escalas: la de GCS y otras de confianza interpersonal, locus de control y autoestima. Al analizar los datos, los autores decidieron eliminar 5 de los 14 ítems de la escala de GCS, dado que su peso factorial era menor de .40⁴². Los ítems conservados fueron el 1, 2, 3, 4, 6, 8, 9, 10 y 11. La fiabilidad de estos nueve ítems (Alfa = .70) fue considerada aceptable por los autores. Adicionalmente, las puntuaciones en GCS mostraron una correlación significativa y negativa con las de las escalas de confianza interpersonal (al igual que en el estudio de Levine y McCornack, 1991), y locus de control, mientras que no se encontró relación con la escala de autoestima (véase Bond y Lee, 2005b).

En el segundo experimento participaron un total de 64 presos como declarantes, y otros 64 como jueces. En primer lugar se administró a los participantes la escala de GCS y la escala de confianza interpersonal empleadas en el estudio anterior. Posteriormente, se indicó a los jueces que su tarea consistía en juzgar la veracidad de seis declaraciones (tres verdades y tres mentiras) que otro preso iba a presentar. Juez y declarante se sentaron cara a cara, aunque se instruyó al juez para que permaneciera en silencio a lo largo de las 6 declaraciones. Bond y Lee (2005b) estiman que en estas condiciones la situación era similar al nivel de baja sospecha estado del estudio de Levine y McCornack (1991). Adicionalmente, se pidió a los jueces que indicasen en una escala tipo *Likert* de cinco puntos lo seguros que estaban de cada juicio de veracidad. En este segundo estudio la escala de GCS tuvo una fiabilidad (Alfa =.77) superior a la del estudio anterior. De nuevo, las puntuaciones de la escala de confianza interpersonal correlacionaron significativa y negativamente con las de la escala de GCS (véase Bond y Lee, 2005b).

Para valorar la utilidad predictiva de la escala de GCS, las puntuaciones se tricotomizaron, creándose un grupo de baja GCS, un grupo de GCS intermedia

⁴² En este sentido, Bond y Lee (2005b) utilizaron el mismo criterio de conservación de ítems que Levine y McCornack (1991).

y un grupo de alta GCS. Los autores emplearon un análisis de regresión de las puntuaciones en GCS sobre los juicios de mentira, hallando que tales puntuaciones no eran una herramienta útil para predecir los juicios. Así pues, concluyeron que aunque la escala de GCS es una medida fiable de la sospecha comunicacional generalizada, tal como se encontró en el primer estudio, su valor predictivo resulta cuestionable.

En definitiva, al inicio de este apartado, en base a los resultados de Masip, Garrido, Herrero, Antón, et al. (2006), se planteaba la existencia de un sesgo confirmatorio en los juicios de los policías. Este sesgo actuaría haciendo que cuando los policías tienen que juzgar la veracidad de una declaración lo hagan tratando de confirmar que ésta es falsa. Para explicar la existencia de este sesgo se ha acudido a los conceptos de sospecha estado y GCS, y se han presentado distintos estudios sobre el efecto de estos tipos de sospecha en los juicios de veracidad. Respecto a la sospecha estado, la investigación muestra que: (a) inducir niveles moderados de sospecha estado puede aumentar la precisión al reducir el sesgo de veracidad que normalmente se observa en situaciones en las que los detectores no tienen motivos para desconfiar de la veracidad de los emisores (McCornack y Levine, 1990); no obstante, (b) los niveles elevados de sospecha estado aumentan la cantidad de juicios de mentira sin mejorar la precisión global (McCornack y Levine, 1990; Millar y Millar, 1997; Stiff et al., 1992; Toris y DePaulo, 1984). Respecto a la CGS, se ha observado que aunque Levine y McCornack (1991) desarrollaron un buen instrumento para medir este tipo de sospecha, no se ha confirmado con claridad que la GCS sea un buen predictor del sesgo hacia la mentira. En el apartado siguiente expondremos cómo entendemos que sospecha estado, GCS y experiencia policial se relacionan para explicar el comportamiento de los policías ante una tarea de detección del engaño.

SOSPECHA ESTADO, EXPERIENCIA POLICIAL, GCS, Y JUICIOS DE MENTIRA DE LOS POLICÍAS

UNA EXPLICACIÓN DE LA TENDENCIA DE LOS POLICÍAS A JUZGAR LAS DECLARACIONES COMO FALSAS BASADA EN LA GCS

En el Anexo F se expone de forma esquemática el proceso a través del cual hipotetizamos que se desarrolla el sesgo de los policías a juzgar las declaraciones como falsas. A continuación se ofrece una explicación detallada.

No cabe duda de que los policías participan a menudo en lo que Robinson (1996) denomina “interacciones competitivas”. A diferencia de las interacciones cooperativas, en que todos los participantes persiguen idénticos fines en favor del bien común, en las interacciones competitivas cada persona persigue su propio interés, a menudo en detrimento de los intereses de los demás. Una interacción abiertamente competitiva sería, por definición, la típica situación generadora de sospecha estado, ya que a las personas implicadas, conscientes de la naturaleza de la situación, les preocuparía la posibilidad de que los demás les engañaran. Para un policía, una entrevista policial es con toda probabilidad una interacción competitiva. El entrevistador a menudo desconfía y se cuestiona lo dicho por el entrevistado, y esto es en cierto modo adaptativo, ya que los policías no pueden creer ciegamente todo lo que les dicen las personas a las que interrogan. Sin embargo, la exposición repetida de los policías a este tipo de situaciones podría dar lugar a una cronificación de la sospecha estado, de modo que los policías acabarán creyendo que las personas normalmente mienten, es decir, desarrollando una GCS.

Si entonces los policías entraran en la situación de evaluación de la credibilidad con la creencia previa de que el testigo va a mentir (alta GCS), cabría la posibilidad de que, al hacer sus juicios, no atendieran o procesaran la información verbal y no-verbal transmitida por el emisor, sino que, guiados por su alta GCS, los policías resolvieran directamente que el testigo miente. Estaríamos hablando de unos juicios de veracidad basados en heurísticos de decisión. Los policías podrían desarrollar un heurístico de decisión que les llevase a concluir que los demás mienten sin procesar la información de forma

paralela, pero inversa, a como Stiff et al. (1992) plantean que se desarrolla el sesgo de veracidad entre los miembros de parejas estables. Según estos autores, a medida que una relación entre dos personas se va desarrollando, va aumentando la confianza entre ellas, así como la necesidad de percibir que la persona con la que se comparte la relación es sincera. Este planteamiento es comprensible, dado que escrutar constantemente lo que la otra persona dice podría crear un clima de desconfianza y tensión que no favorecería el desarrollo de la relación, mientras que simplemente aceptar como cierto lo que nos dice nuestra pareja se interpreta como signo de confianza mutua, lo que favorece la relación.

Pues bien, es posible que una vez que los policías (siguiendo el proceso antes detallado), han desarrollado una elevada GCS, simplemente tiendan a juzgar las declaraciones de los demás como falsas sin procesar la información que reciben. En este caso, dada la ausencia de procesamiento, los policías deberían tomar sus decisiones respecto a la veracidad de las declaraciones con anterioridad a otras personas de las que se esperase que basasen sus juicios en un procesamiento controlado de la conducta del emisor. No obstante, cuando Masip et al. (2003c) compararon el momento en el que un grupo de policías y otro de estudiantes tomaban sus decisiones sobre la veracidad de un conjunto de declaraciones, encontraron que los primeros tendían más a emitir sus juicios hacia el final de la declaración, mientras que los estudiantes tendían más a hacerlo al principio. Adicionalmente, otros trabajos han mostrado que el número medio de indicios conductuales mencionados por los policías no es inferior que el mencionado por los no policías (Masip, 2002; Masip, Garrido y Rojas, 2000). Estos resultados no apoyan la hipótesis de que los policías no procesen la información, y sí los planteamientos de Garrido et al. (2000) y Masip, Garrido, Herrero, Antón, et al. (2006) de que los policías sí procesan la información, pero de forma sesgada.

Desechada, pues, la hipótesis del no procesamiento de la información, y en base a los datos de los estudios que se acaban de presentar, una posibilidad alternativa sería que los policías, aún mostrando una elevada GCS, sin duda atendieran a la información transmitida por el emisor y la procesaran, pero no

de un modo objetivo para averiguar si miente o dice la verdad, sino de forma sesgada para confirmar que sin duda está mintiendo. En definitiva, lo que aquí se plantea es que el proceso de socialización en el Cuerpo Nacional de Policía, en el que los agentes se ven sometidos casi diariamente a situaciones de elevada sospecha estado, genera en ellos una sospecha comunicacional generalizada (GCS) que les lleva a procesar la información procedente del emisor de una forma sesgada para confirmar que, efectivamente, está mintiendo. Esto explicaría por qué, tal como Meissner y Kassin (2002) planteaban al proponer efecto del sesgo del investigador, la experiencia como profesional aumenta la tendencia a hacer juicios de mentira en vez de mejorar la precisión en situaciones de evaluación de la credibilidad.

Prueba empírica del planteamiento anterior

Nuestra investigación respecto al efecto de la experiencia como profesional en la detección del engaño apunta a dos objetivos:

- a. Comprobar si la experiencia dentro del Cuerpo Nacional de Policía incrementa los niveles de GCS.
- b. Examinar si este incremento da cuenta del sesgo de los policías a juzgar las declaraciones como falsas; para ello es necesario calcular las correlaciones entre los niveles de GCS y la tendencia a efectuar juicios de mentira en una tarea de evaluación de la credibilidad.

Estos dos objetivos se alcanzarían mediante el desarrollo de los tres estudios que se exponen a continuación.

Estudio 1

Objetivo

El objetivo de este primer estudio era adaptar el cuestionario de GCS de Levine y McCornack (1991) para poder utilizarlo con muestras de personas españolas.

Participantes y Procedimiento

El cuestionario inicial de 14 ítems de sospecha comunicacional generalizada de Levine y McCornack (1991) se tradujo al español (ver Anexo G) y se aplicó a una muestra de 152 estudiantes de tercer curso de psicología de la Universidad de Salamanca (129 mujeres y 23 hombres, edad: $M = 22.14$ años), que respondieron expresando su grado de acuerdo o desacuerdo con cada ítem en una escala tipo *Likert* que iba de 1 (*fuertemente en desacuerdo*) a 7 (*fuertemente de acuerdo*). A la hora de adaptar este instrumento a nuestro contexto se intentó (al igual que Levine y McCornack con sus muestras) obtener un instrumento unidimensional. Se calculó, pues, un análisis factorial de componentes principales pidiendo al programa informático una solución unifactorial. Se decidió además retener sólo aquellos ítems con un peso factorial mínimo de .40, criterio que también habían seguido Levine y McCornack. Asimismo, se calculó el Alfa de Cronbach para aquellos conjuntos de ítems que pudieran conformar la escala, y se decidió suprimir los ítems cuya eliminación contribuyera al incremento de la fiabilidad del instrumento.

Resultados

Como se muestra en la Tabla 27, los ítems 1, 6 y 11 mostraron inicialmente un peso factorial inferior a .40. Dado que se observó que la fiabilidad de la escala aumentaba eliminando el ítem 1, se decidió repetir los análisis sin dicho ítem. Se observó entonces que el ítem 6 tenía un peso factorial muy bajo (.41), y demás su eliminación suponía un leve aumento en la fiabilidad de la escala, con lo que se repitieron una vez más los análisis suprimiéndolo. Sin los ítems 1 y 6, el ítem 11 seguía mostrando un peso factorial inferior al mínimo establecido, y se observó un aumento en la fiabilidad de la escala si era eliminado. Procedimos, pues, a eliminarlo.

El instrumento final contenía un total de 11 ítems con un peso factorial mínimo de .512 y era unifactorial. El Alfa de Cronbach fue de .82, satisfactorio por lo tanto, y superior a los valores obtenidos por Levine y McCornack (1991) (.71 en el primer estudio y .75 en el tercer estudio) y por Bond y Lee (2005b) (.75).

Tabla 27. *Fiabilidad de la Escala de CGS y Proceso de Eliminación de Ítems.*

Ítem	M del ítem	DE	Todos los ítems		Sin el ítem 1		Sin los ítems 1 y 6		Sin los ítems 1, 6 y 11	
			Peso Factorial	Alfa sin el ítem	Peso Factorial	Alfa sin el ítem	Peso Factorial	Alfa sin el ítem	Peso Factorial	Alfa sin el ítem
1	4.75	1.62	<u>.351</u>	<u>.8212</u>	-	-	-	-	-	-
2	3.52	1.51	.677	.7990	.683	.8007	.700	.7982	.704	.7974
3	3.34	1.41	.619	.8033	.623	.8056	.636	.8044	.640	.8043
4	2.67	1.34	.649	.8022	.649	.8049	.668	.8028	.699	.8028
5	3.15	1.56	.591	.8051	.595	.8075	.611	.8060	.610	.8065
6	2.97	1.52	<u>.399</u>	.8188	<u>.405</u>	<u>.8221</u>	-	-	-	-
7	3.72	1.52	.572	.8052	.570	.8085	.575	.8086	.575	.8090
8	2.98	1.13	.588	.8050	.585	.8085	.590	.8086	.580	.8104
9	3.68	1.30	.602	.8048	.601	.8076	.588	.8095	.602	.8087
10	3.06	1.78	.533	.8091	.532	.8129	.526	.8155	.512	.8185
11	1.75	1.12	<u>.342</u>	.8177	<u>.337</u>	.8215	<u>.344</u>	<u>.8223</u>	-	-
12	2.40	1.00	.685	.8023	.692	.8037	.677	.8048	.677	.8049
13	3.28	1.24	.511	.8113	.522	.8133	.512	.8151	.526	.8148
14	3.76	1.27	.589	.8051	.584	.8085	.572	.8101	.582	.8099
Alfa de Cronbach			.8193		.8218		.8221		<u>.8223</u>	

Calculamos también un análisis factorial confirmatorio con el programa LISREL 8.5, introduciendo los 11 ítems que se habían retenido. Tal como se muestra en la Tabla 28, el análisis mostró un ajuste satisfactorio. Por lo tanto, el conjunto final de ítems tenía una buena fiabilidad y una sólida estructura unidimensional.

Tabla 28. *Índices de Ajuste del Modelo a los Datos.*

GFI	0.90
NFI	0.89
NNFI	0.93
CFI	0.94
AGFI	0.86
RMR	0.06

Conclusiones

En este primer estudio se consiguió adaptar con éxito el cuestionario de Levine y McCornack (1991) para utilizarlo con muestras españolas. Los datos respecto a la fiabilidad de la herramienta (Alfa = .82) y los resultados del análisis confirmatorio (Tabla 28), indican que los 11 ítems que finalmente se conservaron constituyen un instrumento adecuado para medir la GCS. La puntuación media de los estudiantes en la escala fue $M = 35.57$, $DE = 9.36$. El

siguiente paso fue administrar el instrumento a policías noveles y veteranos para comprobar si la experiencia dentro del Cuerpo Nacional de Policía se relaciona con las puntuaciones en GCS.

Estudio 2

Objetivo e Hipótesis

El objetivo era analizar la relación entre la experiencia como policía y la GCS. Hipotetizamos que las puntuaciones en la escala de GCS de los policías con dilatada experiencia profesional serían significativamente mayores que las de los policías que acababan de ingresar en el cuerpo y que las de no policías (estudiantes de psicología).

Participantes

La muestra de estudiantes fue la del estudio anterior, puesto que ya disponíamos de sus puntuaciones en la escala de GCS. Se tomaron además dos muestras de policías, constituidas por alumnos de la Escala Ejecutiva del Centro de Formación de la Policía de Ávila. La muestra novel ($n = 89$) estaba compuesta por alumnos de segundo curso que habían accedido a la Escala Ejecutiva mediante el procedimiento de oposición libre. Llevaban por lo tanto menos de dos años en el Cuerpo Nacional de Policía. La muestra con experiencia ($n = 88$) la constituían policías que habían accedido a la Escala Ejecutiva mediante la vía de promoción interna. Su experiencia profesional media era de 18.10 años⁴³. Si bien la mayoría de los estudiantes eran mujeres (129 en comparación con solamente 23 hombres), la mayoría de los policías noveles eran hombres (55 hombres frente a 16 mujeres; 18 participantes de esta muestra no indicaron su sexo) y todos los policías con experiencia que indicaron su sexo (79 de los 88 de la muestra) eran hombres. Estas diferencias entre los grupos fueron, por supuesto, significativas, $\chi^2_{(2)} = 174.14$, $p < .001$.

⁴³ En el Centro de Formación de la Policía de Ávila se imparte la formación que reciben los alumnos de nuevo ingreso para acceder a la escala Básica (categoría de Policía) y la que reciben los alumnos para acceder (nuevo ingreso) o ascender (promoción interna) a la Escala Ejecutiva. En el Anexo H se ofrece información más detallada sobre las Escalas, Categorías, y forma de acceso a cada una. No obstante, el lector interesado, puede acudir a la página web de la Policía Nacional ("Centro de promoción", sin fechar, <http://www.policia.es/dfp/index.htm>).

Las tres muestras también se diferenciaron en la edad, $F_{(2,328)} = 8.86, p < .001$; el grupo de estudiantes fue el más joven ($M = 22.14$ años), seguido del grupo de policías noveles ($M = 28.32$ años), y del grupo de policías expertos ($M = 41.49$ años). Las pruebas *post-hoc* realizadas indicaron que los tres grupos se diferenciaban significativamente.

Procedimiento

El cuestionario de GCS se administró a las muestras policiales durante sus clases de psicología en el Centro de Formación. En los análisis, los datos de estas dos muestras de policía se compararían también con las puntuaciones de los 152 estudiantes de psicología utilizados en el Estudio 1 como grupo normativo para la adaptación de la escala. El grupo de estudiantes se consideró como grupo control. Una vez que los participantes hubieron completado el cuestionario, éste se recogió y se proporcionó información sobre el estudio.

Resultados

La puntuación media global obtenida en la escala de GCS fue 35.94 ($DE = 10.29$). La puntuación media de los estudiantes fue 35.57 ($DE = 9.36$), la de los policías noveles 33.13 ($DE = 10.11$) y la de los policías con experiencia 39.42 ($DE = 11.09$). Se calculó en primer lugar un ANCOVA a través del Modelo Lineal General del SPSS. El grupo de pertenencia (estudiantes, policías noveles y policías veteranos) de los participantes, se introdujo como variable independiente, el sexo y la edad de los participantes se introdujeron como covariables, y las puntuaciones en GCS como variable dependiente. Los análisis mostraron que ninguna de las covariables tuvo efectos significativos sobre las puntuaciones en GCS⁴⁴. Por su parte, el efecto del grupo de

⁴⁴ Al describir las características de las tres muestras, se había indicado que existían diferencias en edad y género entre ellas. Dichas diferencias eran inevitables, y la única opción era controlarlas estadísticamente. Hasta hace unos pocos años, las mujeres no podían acceder a la policía (razón por la que en nuestra muestra de policías con experiencia no había mujeres), e incluso actualmente la profesión de policía puede seguir siendo trabajo típicamente masculino (lo que explica que en el grupo de policías noveles hubiera más hombres que mujeres). Por el contrario, la proporción de mujeres que estudian psicología es mucho mayor que la de hombres. Respecto a la edad, los estudiantes eran más jóvenes que los policías

pertenencia fue significativo, $F_{(2,301)} = 4.42$, $p = .013$. Se calculó entonces un ANOVA unifactorial sin las covariables⁴⁵. Se observó un efecto significativo de la variable grupo de pertenencia sobre las puntuaciones en GCS, $F_{(2,328)} = 8.86$, $p < .001$, (véase la Figura 1).

Una prueba *post-hoc* de *Scheffe* mostró que la puntuación de los estudiantes universitarios ($M = 35.57$, $DE = 9.36$), sin experiencia policial en absoluto, no difería significativamente de la de los policías noveles ($M = 33.13$, $DE = 10.11$), con poco más de un año de experiencia policial, $z = 2.44$, $p = .193$; es decir, ambos grupos tenían niveles similares de GCS. Por el contrario, los policías con experiencia, los cuales habían experimentado un prolongado proceso de socialización profesional dentro del cuerpo, mostraron una GCS significativamente superior ($M = 39.42$, $DE = 11.09$) a las de los estudiantes universitarios, $z = 3.85$, $p = .018$, y los policías noveles, $z = 6.29$, $p < .001$.

noveles porque los primeros todavía estaban en la universidad, mientras que haber obtenido un título universitario es uno de los requisitos para entrar en la Escala Ejecutiva (ver Anexo H). Por otro lado, es evidente que adquirir experiencia requiere el paso del tiempo, lo que explica las diferencias de edad entre el grupo de policías noveles y con experiencia. Además de haber realizado un ANCOVA con el género y la edad como covariables, se realizaron análisis adicionales para examinar los posibles efectos de dichas variables. En primer lugar, se comprobó que hombres y mujeres no se diferenciaban significativamente en sus puntuaciones en GCS ni cuando se consideraron conjuntamente todos los datos de las tres muestras, $t_{(300)} = 1.68$, $p = .095$, ni cuando se consideraron por separado los estudiantes, $t_{(150)} = 0.83$, $p = .410$, y los policías noveles, $t_{(69)} = 0.37$, $p = .711$. Dado que todos los policías con experiencia eran hombres, estos análisis no se realizaron con este grupo; sin embargo, considerando conjuntamente policías noveles y con experiencia, tampoco se encontraron efectos del sexo sobre las puntuaciones en GCS, $t_{(148)} = 1.12$, $p = .266$. Respecto a la edad, se encontró que correlacionaba significativamente con las puntuaciones en GCS cuando se analizaron los datos de las tres muestras conjuntamente, $r(N = 329) = .17$, $p = .002$. No obstante, este dato no debería ser sorprendente, dada la necesidad de que pase el tiempo para adquirir experiencia y la evidencia de que los policías más expertos mostraron las más altas puntuaciones en GCS (ver el texto). Sin embargo, para considerar a la edad responsable de las puntuaciones en GCS, la correlación edad – GCS debía mantenerse dentro de los diferentes grupos tomados aisladamente. Se realizaron análisis de correlación que indicaron que este no era el caso ni entre los estudiantes, $r(N = 152) = .00$, $p = .982$, ni entre los policías noveles, $r(N = 89) = .01$, $p = .940$, ni entre los policías con experiencia, $r(N = 88) = .11$, $p = .327$. En conclusión, parece que las variables demográficas como tales no tuvieron un efecto directo sobre las puntuaciones en GCS.

⁴⁵ Una vez establecido que las características demográficas de los participantes no afectaban a la variable dependiente, podía calcularse un ANOVA sin ellas. Esto permitiría incluir en el análisis los datos de los 27 participantes (18 policías noveles y 9 con experiencia) excluidos al realizar el ANCOVA por no haber indicado el género.

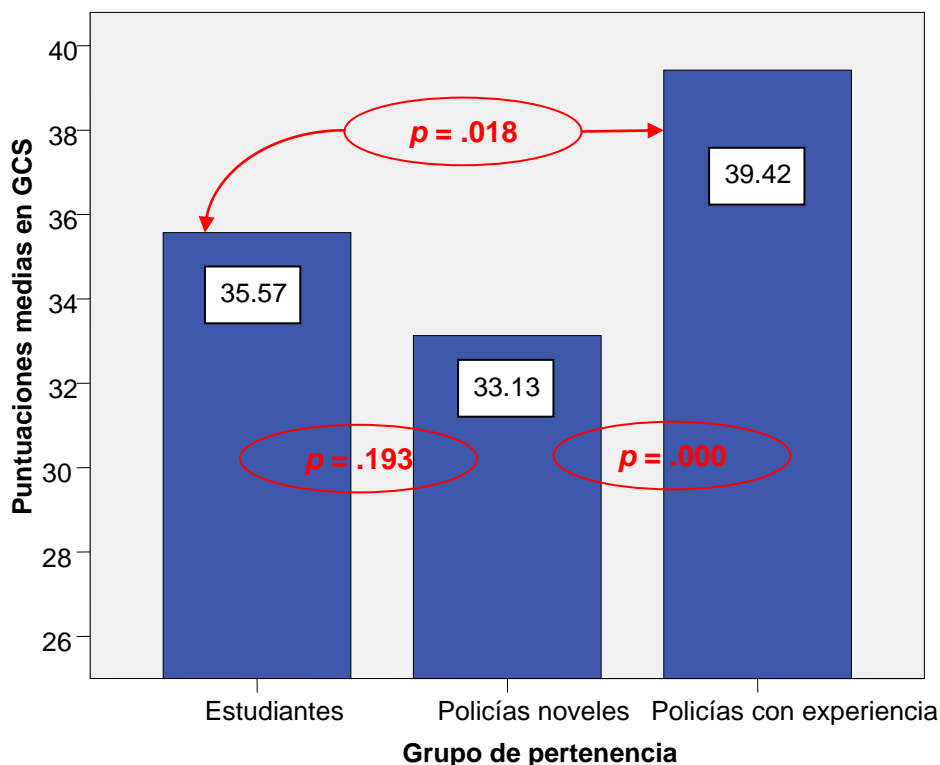


Figura 1. Puntuaciones Medias en GCS de Estudiantes, Policías Noveles y Policías con Experiencia.

Conclusiones

Los resultados de este estudio indican que, tal como se había hipotetizado, la experiencia dentro del Cuerpo Nacional de Policía incrementa la sospecha comunicacional generalizada de estos profesionales. Los presentes datos, además, descartan la posibilidad de que las personas originariamente más desconfiadas o recelosas opten y sean seleccionadas para acceder a la policía. Si éste fuera el caso, los policías noveles habrían obtenido puntuaciones equivalentes a las de la muestra con experiencia, en lugar de obtener puntuaciones equivalentes a las del grupo de estudiantes. O por lo menos sus puntuaciones hubieran superado a las de este último grupo. Por lo tanto, lo que incrementó la GCS de los participantes fue la experiencia como policía. Este incremento, no obstante, no es súbito, sino gradual; parece suceder a lo largo de los años. Sin duda, los policías noveles “eran” (se consideraban a sí mismos) tan policías como los que tenían experiencia; algo que resulta razonable en una institución como la Academia de Policía, caracterizada por

una fuerte presión hacia la uniformidad, un fuerte sentimiento de pertenencia grupal entre los policías y una disciplina cuasi-militar. Sin embargo, los policías noveles puntuaban como los estudiantes, no como sus colegas más experimentados. Por lo tanto, probablemente sea la implicación diaria de los policías en situaciones competitivas la que, a largo plazo, convierte a los agentes en personas desconfiadas, dando lugar a una alta GCS entre ellos.

Esta alta GCS podría, en principio, explicar el efecto del sesgo del investigador descrito por Meissner y Kassin (2002). Si los policías entran en la situación de evaluación de la credibilidad con la sospecha general de que el testigo está mintiendo (alta GCS), entonces es esperable que, tal como Masip, Garrido, Herrero, Antón, et al. (2006) encontraron, atiendan selectivamente a claves asociadas con juicios de mentira, interpreten los indicadores ambiguos como claves del engaño, y no tengan en cuenta aquellas conductas asociadas con juicios de verdad. Esto, en último término, dará lugar a juicios de mentira.

Un aspecto importante de los datos aquí descritos concierne a la representatividad de una y otra muestra policial. La muestra de policías noveles tenía una serie de características que la hacían distinta del conjunto global de policías: todos sus integrantes eran jóvenes, habían alcanzado licenciaturas o, por lo menos, diplomaturas universitarias, y su experiencia policial era casi nula. Por el contrario, la edad, nivel de estudios y experiencia de la muestra experimentada no eran sólo más heterogéneas, sino además probablemente más representativas de la población policial. Dado que esta muestra con experiencia tenía niveles de GCS particularmente elevados, probablemente se pueda afirmar que la policía en su conjunto presenta niveles altos de GCS, con lo que esto implica en relación con no prestar credibilidad a las declaraciones exculpatorias de sospechosos que en realidad pueden ser inocentes.

Hasta el momento, este estudio solamente ha permitido constatar los mayores niveles de GCS de los policías con más experiencia, lo cual supone el primero de los dos objetivos anteriormente propuestos. El siguiente paso, definido por el segundo objetivo, sería examinar si este incremento en la GCS da cuenta del sesgo de los policías a juzgar las declaraciones como falsas; para ello es necesario calcular si existe una relación significativa entre la GCS

de los policías y su tendencia a efectuar juicios de mentira en una tarea de evaluación de la credibilidad. Desgraciadamente, no fue posible someter a los policías que habían cumplimentado el cuestionario de GCS a dicha tarea. Esto obedeció a razones relacionadas con los horarios y el funcionamiento interno del Centro de Formación de la Policía de Ávila, en donde se habían recogido los datos de las muestras policiales. Se hizo necesario, por lo tanto, aplicar el cuestionario de GCS a nuevas muestras policiales para, más adelante, someterlas a la tarea de evaluación de la credibilidad. Este trabajo se describe a continuación.

Estudio 3

Objetivo

El objetivo de este estudio era analizar la relación entre la experiencia como policía, la GCS y los juicios de mentira. Pretendíamos también replicar los resultados del estudio anterior sobre la GCS y los hallazgos de la investigación que muestran una mayor tendencia a hacer juicios de mentira entre los policías que entre los no policías. Adicionalmente, dado que la investigación ha encontrado que los policías muestran un nivel de confianza en sus propios juicios superior a los no profesionales, se decidió comprobar si esto también sucedía con nuestras muestras de policías y estudiantes, y si policías nuevos y con experiencia diferían en confianza. Por último quisimos ver si la confianza de los policías en sus juicios se relacionaba de alguna forma con sus niveles de GCS.

Para examinar estas cuestiones debíamos disponer de muestras de no-policías y de policías noveles y veteranos que: (a) cumplimentaran el cuestionario de GCS y (b) de forma independiente y sin asociar ambas tareas, visualizaran los vídeos e indicaran qué declaraciones eran verdaderas y cuáles falsas. Por último, una vez obtenidos estos datos habría que examinar las correlaciones entre las puntuaciones de GCS y la proporción de declaraciones consideradas falsas por los participantes.

Materiales

Se empleó la escala de GCS del Estudio 2. Además, para la tarea de evaluación de la credibilidad se usaron unas grabaciones audiovisuales elaboradas y utilizadas anteriormente por Masip (2002) y Masip, Garrido y Herrero (2006, 2007). Se trata de un conjunto de cuatro vídeos, etiquetados A1, A2, B1 y B2, en cada uno de los cuales aparecen 12 personas (emisores) respondiendo a tres preguntas sobre un robo que han presenciado en vídeo. La mitad de los emisores miente y la otra mitad dice la verdad. Al elaborar estos *clips* se contrabalancearon una serie de variables, tales como la versión del vídeo del robo que habían visto los emisores (seis de ellos [tres sinceros y tres mentirosos] habían visto una versión y los otros seis otra versión distinta), el orden en que se habían formulado las tres preguntas y el sexo del emisor. Las muestras de estudiantes, policías noveles y policías veteranos se dividirían en cuatro subgrupos, cada uno de los cuales juzgaría las declaraciones contenidas en uno de los vídeos.

Participantes

Un subconjunto de los estudiantes que habían cumplimentado el cuestionario de GCS hizo además, como ejercicio aparte, una serie de juicios de credibilidad. En cuanto a los policías, hubo que utilizar muestras nuevas de alumnos de la Escala Ejecutiva del Centro de Formación de Ávila que pudieran hacer ambas tareas. Doscientos veintisiete policías cumplimentaron el cuestionario de GCS y 302 hicieron el ejercicio de evaluación de la credibilidad. La mayoría había hecho ambas tareas, pero era importante poder emparejar los dos cuestionarios (la escala de GCS y aquel en el que habían expresado sus juicios de veracidad). Para ello, se tuvo en cuenta su sección (había cerca de 40 alumnos por sección) y su fecha de nacimiento (que debían indicar en ambos cuestionarios). Sin embargo, muchos policías no señalaron su fecha de nacimiento, lo cual hizo imposible emparejar un gran número de cuestionarios. La muestra se redujo todavía más al excluir los cuestionarios en los que no figuraba algún dato demográfico (sexo o edad) o la cantidad de experiencia policial, así como aquellos en los que no se habían respondido todas las preguntas cruciales (es decir, se habían dejado ítems de la escala de GCS en

blanco o no se habían hecho los 12 juicios de credibilidad en la tarea posterior). Se excluyeron, además, los datos de un pequeño grupo de policías que tenía más experiencia que los nuevos pero menos que los veteranos. Se trataba de los que habían accedido originalmente al Cuerpo a través de la Escala Básica, pero que luego habían accedido a la Escala Ejecutiva a través del Turno Libre y no a través de la promoción interna.

Este proceso de exclusión dejó excesivamente descompensadas, en términos de número de participantes, las celdas constituidas por el cruce entre población (estudiantes, policías noveles, policías veteranos) y vídeo visualizado durante la tarea de evaluación de la veracidad (A1, A2, B1, B2). Por ello, se decidió igualar la muestra de todas las celdas al tamaño de aquella en la que hubiera menos participantes (en concreto, 18). Se eliminaron, pues, los participantes sobrantes al azar, pero con el condicionante de que la edad de los miembros de cada grupo no difiriera entre los vídeos-estímulo. La muestra final constó de 72 estudiantes, 72 policías noveles y 72 policías veteranos. En cada grupo, 18 personas habían juzgado las declaraciones de cada vídeo.

Procedimiento

Las dos tareas (cumplimentar el cuestionario de GCS y evaluar la credibilidad de 12 declaraciones en vídeo) se hicieron en momentos distintos del curso, y se tomaron precauciones para que los participantes no las relacionaran. Ambas se hicieron en grupo, para lo cual el experimentador acudió al aula que ocupaban de ordinario los alumnos. Las instrucciones dadas a los policías antes de la administración del cuestionario de GCS figuran en el Anexo I y las correspondientes a la tarea de evaluación de la credibilidad se recogen en el Anexo J. Nótese que, además de juzgar la veracidad, los participantes tuvieron que indicar la seguridad o confianza en cada uno de sus juicios en una escala que abarcaba de 1 a 7 en la cual mayores puntuaciones significaban más seguridad.

Resultados

a) Sospecha Comunicacional Generalizada

El valor promedio global en el cuestionario de GCS fue $M = 35.88$, ($DE = 10.30$). Los estudiantes (ver Figura 2) puntuaron 34.94, ($DE = 9.68$) los policías noveles 33.90, ($DE = 9.70$) y los policías con experiencia 38.81, ($DE = 10.95$). Estos valores son sorprendentemente similares a los obtenidos en el estudio inicial descrito anteriormente. Un ANOVA de un factor mostró que el efecto del grupo de pertenencia fue significativo, $F_{(2, 213)} = 4.69$, $p = .010$, $\eta^2 = .042$. Una prueba *post-hoc* de *Scheffe* mostró que las diferencias entre estudiantes y policías noveles no fueron significativas ($p = .827$), pero sí lo fueron las diferencias entre policías con experiencia y policías noveles ($p = .016$) y, de forma marginalmente significativa, las diferencias entre policías con experiencia y estudiantes ($p = .075$). Así pues, de forma general, estos resultados suponen una réplica de lo encontrado en el estudio anterior.

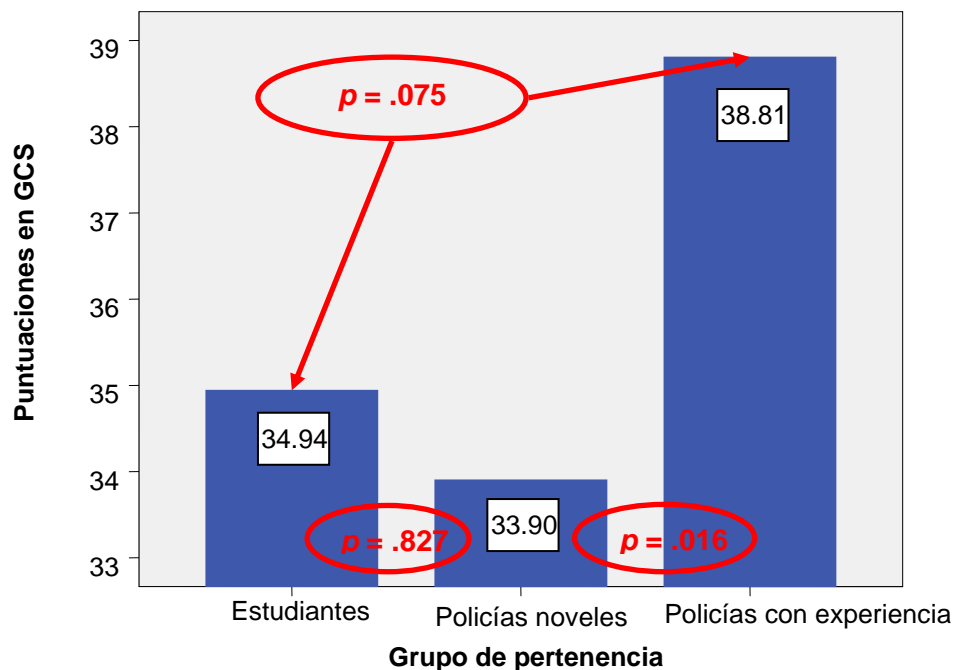


Figura 2. Puntuaciones Medias de GCS de Estudiantes, Policías Noveles y Policías con Experiencia.

b) Precisión, porcentaje de juicios de verdad y sesgos

Se calcularon dos Análisis de Varianza (ANOVAS) mixtos con el grupo de pertenencia (estudiantes vs. policías noveles vs. policías con experiencia) como variable independiente inter-sujeto y el valor de verdad (si las declaraciones eran verdaderas o falsas) como variable independiente de medidas repetidas. Las variables dependientes fueron la precisión de los juicios (porcentaje de aciertos) y el porcentaje de juicios de mentira (PJM).

El ANOVA sobre la precisión no mostró ningún efecto significativo de la variable grupo de pertenencia, $F_{(2, 213)} = 1.85$, $p = .308$, $\eta^2 = .011$, es decir, la precisión de los estudiantes ($M = 54.86$, $DE = 12.80$), los policías noveles ($M = 57.18$, $DE = 12.22$) y los policías con experiencia ($M = 54.05$, $DE = 12.89$), fue prácticamente la misma y cercana a los niveles del azar, encontrándose dentro del rango de valores normalmente hallado en la investigación en este campo.

Sí hubo un efecto significativo de la variable valor de verdad, indicativo de que, a nivel global, las declaraciones falsas ($M = 57.64$, $DE = 18.32$) se identificaron en mayor medida que las declaraciones verdaderas ($M = 53.09$, $DE = 21.16$), $F_{(1, 213)} = 5.19$, $p = .024$, $\eta^2 = .024$. Sin embargo, esta relación se veía moderada por el grupo de pertenencia (ver Figura 3), ya que la interacción entre valor de verdad y grupo de pertenencia fue significativa, $F_{(2, 213)} = 8.98$, $p < .001$, $\eta^2 = .078$. Esta interacción refleja que si bien tanto los policías con experiencia (las medias para las declaraciones falsas y verdaderas fueron, $M = 60.65$, $DE = 17.54$ y $M = 47.45$, $DE = 20.68$, respectivamente) como los policías noveles ($M = 60.88$, $DE = 17.54$ y $M = 53.47$, $DE = 19.16$) sí que acertaron más las declaraciones falsas que las verdaderas (respectivamente, $F_{(1, 123)} = 14.54$, $p < .001$, $\eta^2 = .064$; $F_{(1, 123)} = 4.58$, $p = .033$, $\eta^2 = .021$), entre los estudiantes la tendencia fue la contraria⁴⁶ ($M = 51.39$, $DE = 18.71$ y $M = 58.33$, $DE = 22.38$; $F_{(1, 123)} = 4.03$, $p = .046$, $\eta^2 = .019$). Se hicieron análisis adicionales que indicaron que los estudiantes mostraron una precisión significativamente superior al nivel de azar al juzgar las declaraciones verdaderas, $t_{(71)} = 3.16$, $p = .002$, pero no al juzgar las falsas, $t_{(71)} = 0.63$, $p = .531$. Con los policías noveles

⁴⁶ En este estudio y en los siguientes, los análisis post hoc de los efectos individuales se hicieron con el comando EMEANS de la sintaxis de SPSS. Para las comparaciones múltiples se hicieron ajustes de Bonferroni.

sucedía justo lo opuesto ($t_{(71)} = 1.53, p = .129$, para las declaraciones verdaderas, y $t_{(71)} = 5.34, p < .001$, para las falsas), al igual que con los policías con experiencia ($t_{(71)} = -1.04, p = .300$, para las declaraciones verdaderas, y $t_{(71)} = 5.15, p < .001$ para las falsas).

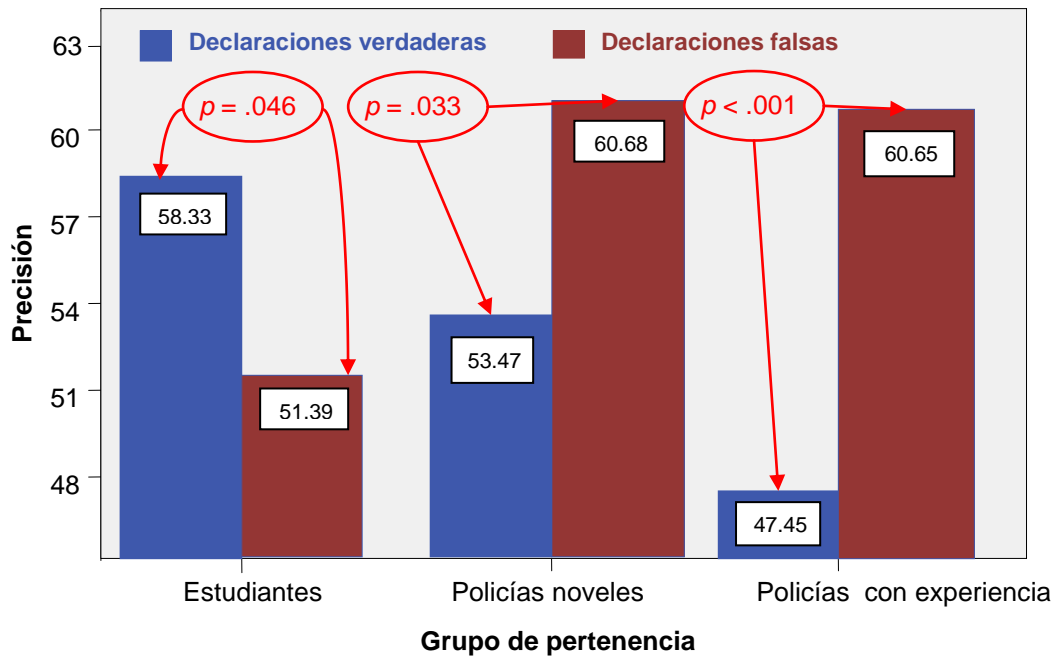


Figura 3. Precisión de los Estudiantes, Policías Noveles y Policías Expertos en Función de la Veracidad de las Declaraciones.

El ANOVA sobre el porcentaje de juicios de mentira (PJM) arrojó un efecto significativo de la variable grupo de pertenencia (ver Figura 4), $F_{(2, 213)} = 8.98, p < .001, \eta^2 = .078$. Las puntuaciones de los policías con experiencia, policías noveles y estudiantes fueron ($M = 56.60, DE = 14.19$; $M = 53.70, DE = 13.55$; y $M = 46.53, DE = 16.18$, respectivamente). Una prueba *post-hoc* de Scheffe indicó que las diferencias entre las puntuaciones de los policías con experiencia y las de los estudiantes fueron estadísticamente significativas ($p < .001$). Las diferencias entre policías noveles y estudiantes fueron también significativas ($p = .015$) Sin embargo, las diferencias entre policías con experiencia y policías noveles no fueron estadísticamente significativas ($p = .498$). Esta tendencia puede dar cuenta de la mayor precisión de los policías al

juzgar declaraciones falsas y de la mayor precisión de los estudiantes al juzgar declaraciones verdaderas.

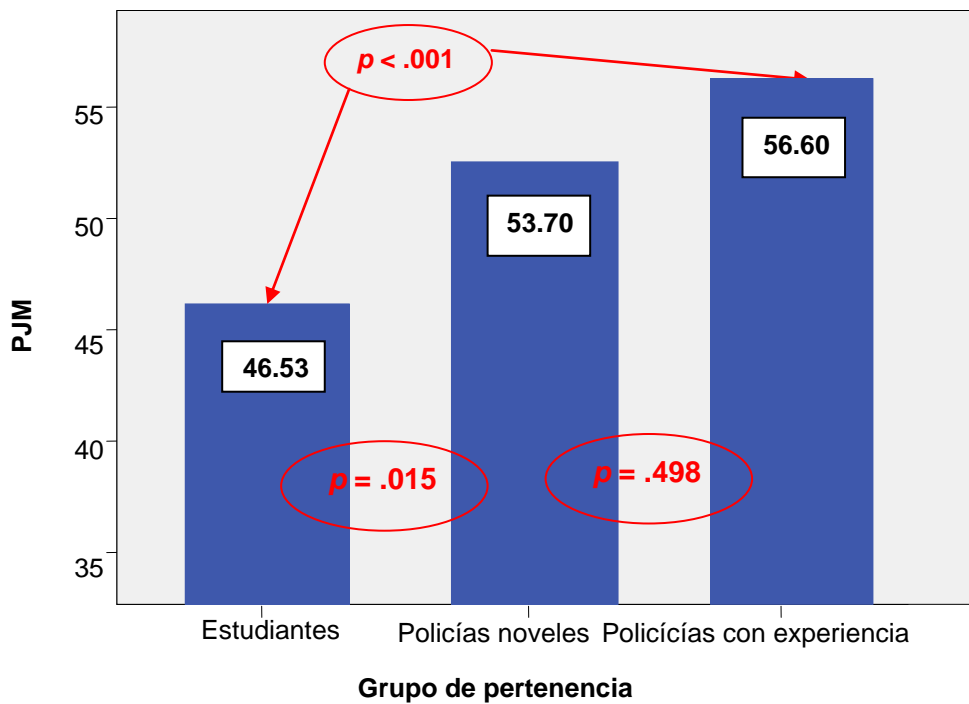


Figura 4. Porcentajes Medios de Juicios de Mentira de los Estudiantes, Policías Noveles y Policías con Experiencia.

El efecto de la variable valor de verdad fue significativo, $F_{(1, 213)} = 38.88$, $p < .001$, $\eta^2 = .154$. Como se puede apreciar en la Figura 5, las declaraciones falsas se juzgaron como tales ($M = 57.64$, $DE = 18.31$) en mayor medida que las verdaderas ($M = 46.91$, $DE = 21.16$), lo que indica que la precisión global, si bien moderada, fue superior al azar.

Los resultados anteriores sugieren que ambos grupos de policías tendían a hacer juicios de mentira, mientras que los estudiantes tendían a hacer juicios de verdad. Para examinar más de cerca estas tendencias, se hicieron algunos análisis sobre B'' . Ésta es una medida derivada de la Teoría de la Detección de Señales (por ej., Green y Swets, 1966; Stanislaw y Todorov, 1999; Swets, Dawes y Monahan, 2000) que refleja el sesgo de respuesta con independencia de la precisión. Se calcula a partir de los Aciertos y de las Falsas Alarmas. Se producen Aciertos cuando una señal (por ej., una mentira) se identifica como

tal. Se producen Falsas Alarmas cuando la ausencia de señal (por ej., una verdad) se juzga como señal (esa verdad se juzga como mentira). Los valores de B'' pueden abarcar desde -1 hasta 1 . Una respuesta sin sesgo arrojaría un valor de $B'' = 0$, los valores negativos reflejan un sesgo a juzgar las declaraciones como falsas y los valores positivos reflejan un sesgo a juzgar las declaraciones como verdaderas.

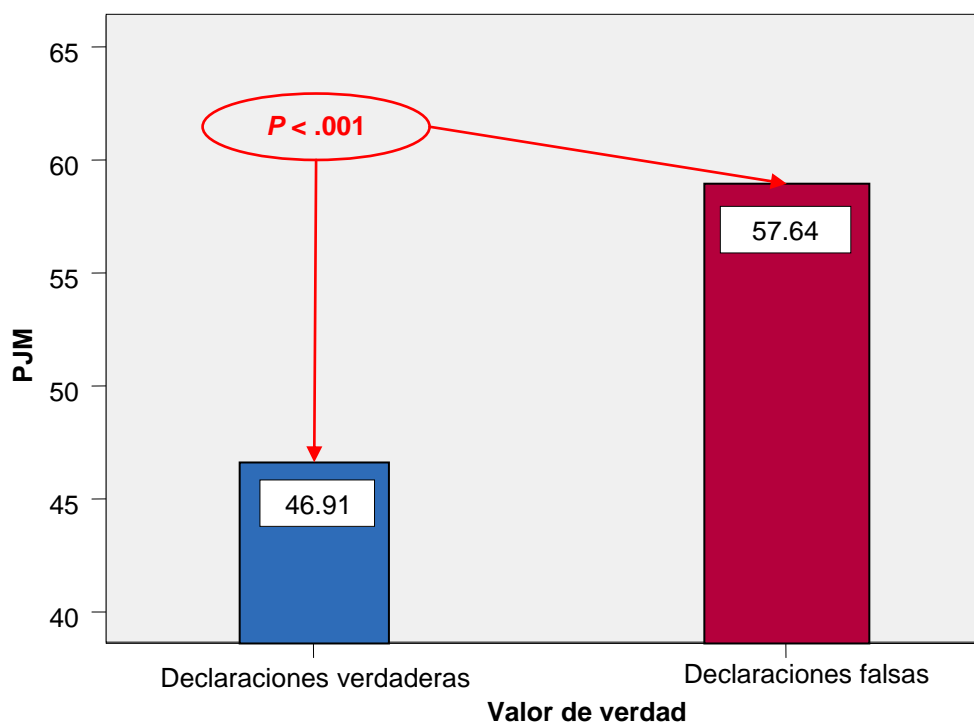


Figura 5. Porcentajes Medios de Juicios de Mentira Emitidos Para Declaraciones Verdaderas y Falsas.

Se realizó un ANOVA de un factor en el que se introdujo como variable independiente el grupo de pertenencia y como variable dependiente las puntuaciones en B'' . El efecto fue significativo, $F_{(2,213)} = 9.56$, $p < .001$, $\eta^2 = .082$. En concreto, los estudiantes mostraban un sesgo hacia la verdad ($B'' = .14$) en sus juicios, mientras que los policías noveles ($B'' = -.14$) y con experiencia ($B'' = -.22$) un sesgo hacia la mentira (ver Figura 6). Una prueba *post-hoc* de *Scheffe* indicó que el sesgo de los policías noveles y con experiencia no se diferenciaba significativamente ($p = .630$), pero las diferencias sí eran significativas entre los estudiantes y estos dos grupos ($p = .006$ y $p < .001$, respectivamente).

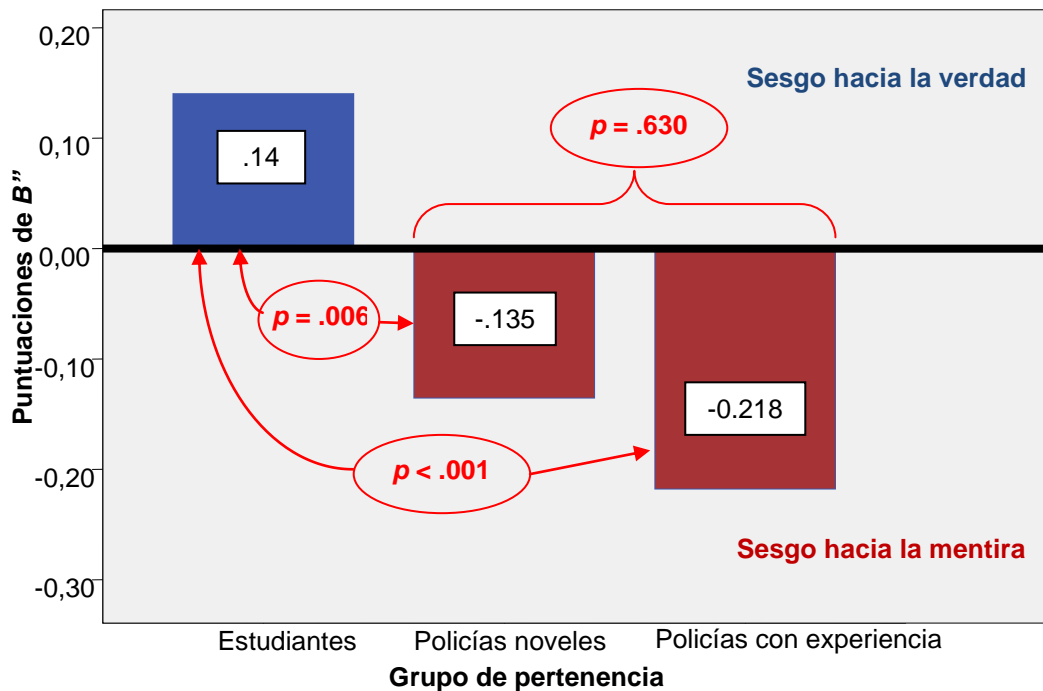


Figura 6. Puntuaciones Medias en B'' de Estudiantes, Policías Noveles y Policías con Experiencia.

Se hicieron análisis adicionales que mostraron que los tres grupos estaban sesgados, ya que en los tres casos fueron significativas las diferencias entre sus valores promedios de B'' y 0. Los estudiantes estaban sesgados hacia la verdad, $t_{(71)} = 2.14$, $p = .036$, mientras que los policías noveles, $t_{(71)} = -2.29$, $p = .025$, y los policías con experiencia, $t_{(71)} = -3.82$, $p < .001$, estaban sesgados hacia la mentira.

En resumen, a diferencia de los estudiantes, que mostraron un sesgo de veracidad, los policías tendieron a juzgar las declaraciones como falsas, y esto fue así con independencia de si tales declaraciones eran realmente falsas o verdaderas.

c) Confianza

Se midió también la confianza en los juicios. Un ANOVA similar a los anteriores en que se introdujo la confianza como variable dependiente mostró un efecto significativo de la variable grupo de pertenencia, $F_{(2, 213)} = 18.18$, $p < .001$, $\eta^2 = .146$. A medida que aumentaba la experiencia, así lo hacía la

confianza (estudiantes: $M = 4.17$, $DE = 0.83$; policías noveles: $M = 4.51$, $DE = 0.75$; policías con experiencia: $M = 4.99$, $DE = 0.86$). Una prueba *post-hoc* de *Scheffe* mostró que las diferencias entre estudiantes y policías con experiencia ($p < .001$), y policías noveles y policías con experiencia ($p = .002$) eran estadísticamente significativas. Sin embargo, las diferencias entre estudiantes y policías noveles ($p = .053$) no alcanzaron significatividad estadística (ver Figura 7).

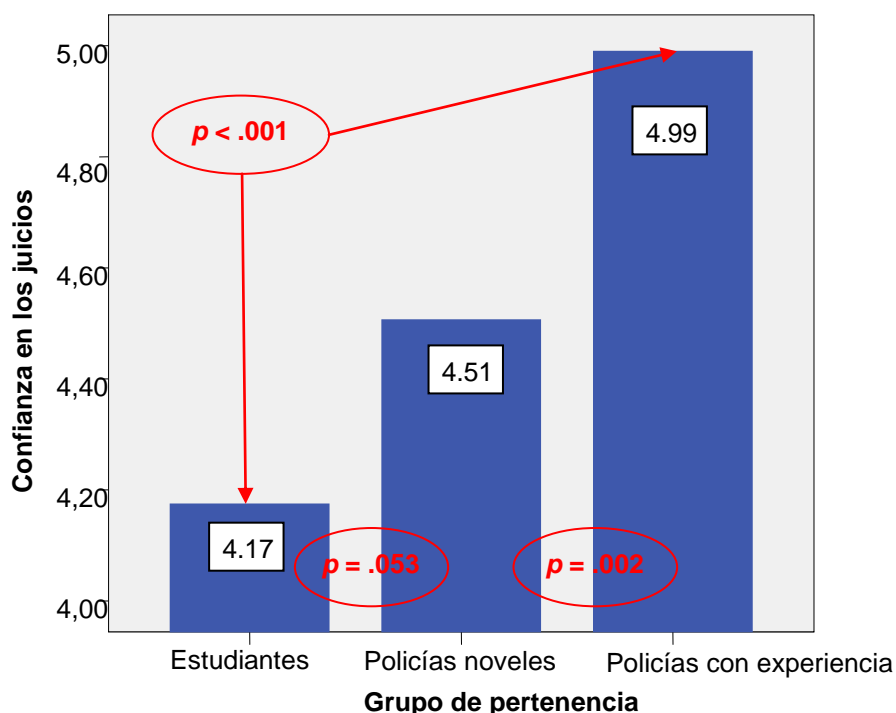


Figura 7. Puntuaciones en Confianza en los Propios Juicios de Estudiantes, Policías Noveles y Policías con Experiencia.

El valor de verdad de la declaración tuvo también un efecto significativo, $F_{(1, 213)} = 6.55$, $p = .011$, $\eta^2 = .030$. Se tuvo más confianza al juzgar las declaraciones falsas ($M = 4.62$, $DE = 0.93$) que al juzgar las verdaderas ($M = 4.50$, $DE = 0.96$) (Figura 8). Aunque la interacción Valor de Verdad X Grupo de pertenencia no mostró un efecto significativo, $F_{(2, 213)} = 1.46$, $p = .234$, $\eta^2 = .014$, calculamos análisis *post-hoc* para examinar si el efecto principal del valor de verdad era debido a los juicios de los policías y no a los de los estudiantes. Se encontró que, efectivamente, mientras que tanto los policías con experiencia $F_{(1, 123)} = 4.48$, $p = .036$, $\eta^2 = .036$, como los policías noveles, $F_{(1, 123)} = 4.98$, $p =$

.027, $\eta^2 = .023$, mostraban más confianza al juzgar mentiras ($M = 5.08$ para los policías con experiencia; $M = 4.60$ para los noveles) que al juzgar verdades ($M = 4.90$; $M = 4.41$, respectivamente), los estudiantes tenían virtualmente la misma confianza al juzgar mentiras ($M = 4.18$) que al juzgar verdades ($M = 4.17$), $F_{(1, 123)} = 0.01$, $p = .934$, $\eta^2 < .001$.

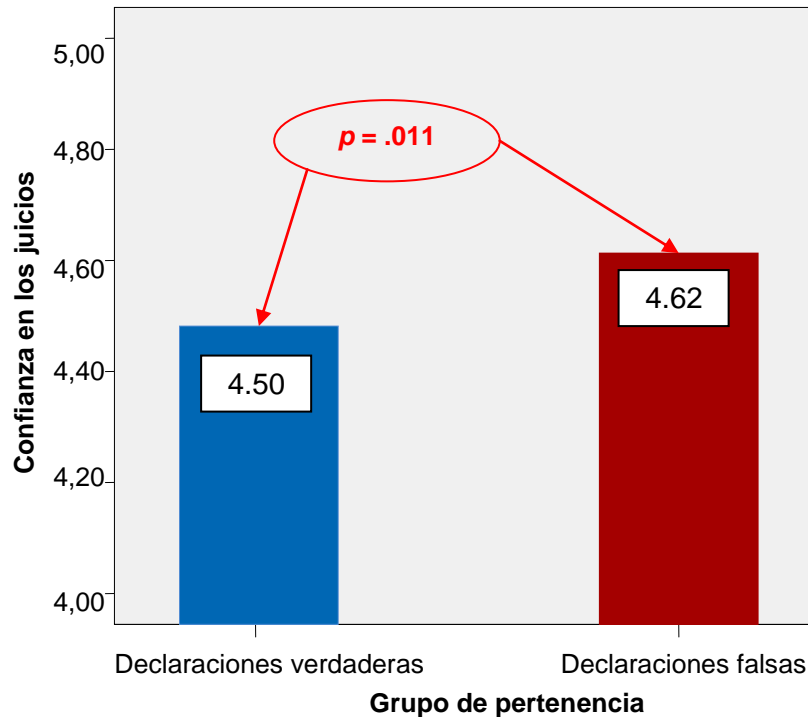


Figura 8. Confianza Para Declaraciones Verdaderas y Falsas.

En línea con lo encontrado en la investigación en este campo, la relación entre confianza y precisión no fue significativa ni a nivel global, ni para declaraciones verdaderas ni para declaraciones falsas⁴⁷.

En resumen: con independencia de si las declaraciones eran verdaderas o falsas, los policías con experiencia mostraron más confianza que los policías noveles. Globalmente la confianza al juzgar declaraciones falsas fue superior

⁴⁷ Cuando se examinaron las correlaciones entre la confianza y la precisión separadamente para las diversas muestras, se encontró que fue significativa entre los estudiantes al juzgar declaraciones verdaderas, $r (N = 72) = .26$, $p = .027$, y entre los policías con experiencia al juzgar declaraciones falsas, $r (N = 72) = .29$, $p = .012$. En ningún otro caso guardó la confianza relación significativa alguna con la precisión.

que la mostrada al juzgar declaraciones verdaderas, aunque esto parece deberse a los policías y no a los estudiantes.

d) Relación entre la GCS y el PJM

Hasta aquí se ha mostrado, entre otras cosas: (a) que los policías hacen más juicios de mentira (y menos juicios de verdad) que los estudiantes, y (b) que la experiencia como policía incrementa la GCS. Falta, en este punto, cerrar el triángulo examinando si la relación entre la GCS y el PJM es significativa y positiva, de modo que a mayores niveles de GCS se produzca un mayor PJM.

Tabla 29. *Correlaciones Entre las Puntuaciones en GCS y el PJM.*

	Declaraciones Verdaderas		Declaraciones Falsas		Global	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Estudiantes	.12	.303	.01	.933	.09	.447
Policías nuevos	.08	.517	-.18	.128	-.06	.613
Policías con experiencia	.01	.951	.00	.991	.01	.969
Total	.10	.147	-.04	.601	.05	.490

Tal como se aprecia en la Tabla 29 el PJM no guarda relación con la GCS. Así pues, el efecto de la GCS sobre el PJM no llegó a ser el que nuestra explicación propuesta predecía.

Adicionalmente se analizó la correlación entre la GCS por un lado y el sesgo de respuesta (*B''*) y las falsas alarmas por otro. Los resultados se presentan en la Tabla 30. Como se puede observar, tampoco se encontró ninguna correlación significativa. Así pues, se puede concluir que la mayor GCS de los policías con experiencia no explica su sesgo a hacer juicios de mentira.

Tabla 30. *Correlaciones Entre las Puntuaciones en GCS y el Sesgo de Respuesta (*B''*).*

	<i>B''</i>		Falsas Alarmas	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Estudiantes	-.10	.416	.12	.303
Policías nuevos	.06	.626	.08	.517
Policías con experiencia	.01	.941	.01	.951
Total	-.04	.539	.10	.147

Conclusiones

En definitiva, si bien se ha mostrado que con la experiencia se incrementa la GCS de los policías, no parece que este incremento tenga un gran impacto

sobre el PJM. De hecho, todos los policías, y no sólo los que tenían experiencia, mostraron ya una acusada tendencia a juzgar las declaraciones como falsas. Sin embargo, sólo los que tenían experiencia presentaron niveles comparativamente elevados de GCS. Debe haber, por consiguiente, otras explicaciones ajenas a la GCS que den cuenta de la tendencia de los policías a juzgar las declaraciones como falsas.

¿POR QUÉ LOS POLICÍAS NOVELES HACEN TANTOS JUICIOS DE MENTIRA COMO LOS VETERANOS?

Una posible explicación de los resultados presentados en el apartado anterior la encontramos en una línea de pensamiento clásica en psicología social y sociología: la que considera que las personas somos “actores” que representamos determinados papeles en función de las circunstancias y de nuestro rol social (por ejemplo, Goffman, 1959). En este sentido, los policías noveles tendrían sin duda sus propios esquemas mentales sobre cómo comportarse cuando, como profesionales, deben juzgar la veracidad de la declaración de un sospechoso (quizá en función del estereotipo de mostrarse “duros”, insistentes, cuestionando o no prestando credibilidad a la declaración, etc.). Además, pueden percibir su actuación como legitimada (por ejemplo, por haber interiorizado que su tarea consiste en descubrir al culpable, en “pillar” al mentiroso). Recuérdese, a este respecto, lo bien que carceleros y presos desempeñaron sus roles en el experimento de la cárcel de Stanford cuando su actuación estaba legitimada (Haney, Banks y Zimbardo, 1973; Zimbardo, Haney, Banks y Jaffe, 1973).

En este sentido, lo que se plantea es que *cuando se asigna el rol social de policías a las personas, éstas empiezan a comportarse tal como creen que debe comportarse un policía en esa situación determinada*, con independencia de que hayan empezado o no a desarrollar determinadas distorsiones de pensamiento, como la creencia de que los demás realmente suelen mentir (la GCS que sí se observó en los policías más veteranos). Es decir, es cuando los policías noveles están realizando una tarea que perciben como “específica del trabajo policial” y “directamente relacionada” con aspectos nucleares de su labor (como juzgar la veracidad de una declaración, o interrogar a un

sospechoso), que entienden que deben comportarse como se espera de ellos (desconfiando, cuestionando, dudando de la veracidad), y es entonces cuando se sienten legitimados. Esto explicaría porqué aunque en los estudios anteriores se observaron diferencias significativas en GCS entre policías noveles y policías con experiencia, cuando estos dos grupos realizaron una misma tarea de evaluación de la credibilidad, propia de su rol como policía⁴⁸, emitieron ambos un porcentaje similar de juicios de mentira, pero significativamente superior al de los estudiantes (ver Figura 4). Es decir, al realizar esta tarea, los policías ejercieron su rol de policía, comportándose tal como interpretaron que debían hacerlo.

La adopción del rol de policía por los alumnos de Turno Libre de la Escala Ejecutiva (sin experiencia policial) puede verse enfatizada por el largo camino necesario para llegar a este punto (la ardua y prolongada preparación de la oposición, las duras pruebas de acceso, etc.), así como por el notable cambio de vida de quienes superan el concurso y las características propias de la socialización en el Centro de Formación (presión hacia la uniformidad, disciplina, contacto con compañeros veteranos, etc.). También por la percepción existente en nuestra sociedad de la policía como “algo aparte” y con unas características distintivas muy acusadas y diferenciadoras del resto de los ciudadanos. Todo esto puede desarrollar la identidad del nuevo policía como tal, generando en él los fenómenos de asimilación endogrupal y contraste con el exogrupo y haciendo que pase a comportarse como él cree que debe hacerlo un policía.

Si esta argumentación es cierta, entonces no se esperarían diferencias entre policías noveles y policías veteranos si, en lugar de cumplimentar la escala de GCS (que no se contextualiza en ninguna situación, sino que mide una tendencia general), cumplimentaran una escala referida a sospecha en situaciones específicamente policiales. Los policías noveles interpretarían esta escala como algo relevante para su labor profesional, y entonces se comportarían mostrando los mismos niveles de desconfianza y sospecha que

⁴⁸ Recuérdese que además, las declaraciones que se juzgaron en la tarea de evaluación de la credibilidad trataban sobre la comisión de un delito, en concreto, el robo de una cartera.

sus compañeros más veteranos. Esta es la cuestión que examinamos en el Estudio 4, que se describe a continuación.

Estudio 4

Objetivo e Hipótesis

En este estudio el objetivo era observar cómo los policías interpretan una situación policialmente relevante en la que, como profesionales, deben juzgar la veracidad de una declaración. Para ello, lo que hicimos fue partir de 13 ítems de la escala de GCS de Levine y McCornack (1991), no referidos a ninguna situación concreta, y elaborar una escala similar⁴⁹ pero contextualizada en una situación policial específica: los interrogatorios (ver Anexo K). Al igual que en la tarea de evaluación de la credibilidad propuesta a los policías en el estudio anterior, en un interrogatorio el policía debe juzgar si el interrogado miente o dice la verdad. En adelante nos referiremos a esta escala como “Escala de Sospecha en Interrogatorios” (ESI).

Si, como hemos planteado, los policías noveles consideran que en este tipo de situaciones deben representar su papel de policía (desconfiando y sospechando de las declaraciones que van a escuchar), se puede hipotetizar que mientras que en la escala de GCS los policías veteranos alcanzarán puntuaciones superiores a las de los policías noveles, en la ESI las puntuaciones de ambos grupos serán similares y elevadas. En otras palabras: esperábamos una interacción Grupo (policías noveles vs. veteranos) x Escala (GCS vs. ESI).

Participantes y Procedimiento

Se tomaron los datos de los 72 policías noveles y los 72 policías con experiencia que habían cumplimentado la escala de GCS en el Estudio 3. Así mismo, se contó con otras dos muestras policiales que deberían cumplimentar

⁴⁹ Se construyó la escala de forma que cada ítem reflejase esencialmente la misma idea que en la escala original de GCS, pero contextualizada en una situación de interrogatorio policial. Esto nos obligó a eliminar el ítem 3 de la escala original, que reza: “la mayoría de la gente sólo te dice lo que ellos creen que tú quieres oír”, pues en un interrogatorio la mayoría de la gente no dice al policía lo que éste espera oír (es decir, no se confiesa culpable), sino que defiende su inocencia.

la ESI: una de 107 policías con experiencia ($M = 23.59$ años de experiencia, $DE = 3.95$), y otra de 74 policías noveles, de los que 56 tenían menos de un año de experiencia y ninguno de los 18 restantes tenía más de tres años de experiencia. Al igual que en los estudios anteriores, los policías de ambas muestras eran alumnos de la Escala Ejecutiva del Centro de Formación de la Policía Nacional de Ávila. La muestra novel estaba compuesta por alumnos de segundo curso, que habían optado a la Escala Ejecutiva mediante el procedimiento de oposición libre, mientras que la muestra con experiencia la constituían policías que habían optado a la Escala Ejecutiva mediante la vía de promoción interna. Ambas muestras completaron la ESI, compuesta inicialmente por 13 ítems, durante el transcurso de sus clases en el Centro de Formación de la Policía.

Resultados

a) Comprobación de la consistencia interna de la ESI

Al igual que se hizo al adaptar el cuestionario original de GCS, se trató de obtener un instrumento unidimensional. Una vez más, se decidió retener sólo aquellos ítems con un peso factorial mínimo de .40. Asimismo, se calcularía el Alfa de Cronbach para aquellos conjuntos de ítems que pudieran conformar la escala, y los ítems cuya eliminación contribuyera al incremento de la fiabilidad del instrumento se suprimirían.

Se calculó un análisis factorial de componentes principales pidiendo al programa informático una solución unifactorial. Teniendo en cuenta los requisitos mencionados en el epígrafe anterior, se eliminaron los ítems 9 y 10, con lo que finalmente se contó con una escala de 11 ítems, todos ellos con un peso factorial superior a .40. El índice de fiabilidad obtenido (Alfa = .73) fue aceptable (ver Tabla 31).

Tabla 31. *Fiabilidad de la ESI y Proceso de Eliminación de Ítems.*

Ítem	M del ítem	DE	Todos los ítems		Sin el ítem 10		Sin los ítems 10 y 9	
			Peso Factorial	Alfa sin el ítem	Peso Factorial	Alfa sin el ítem	Peso Factorial	Alfa sin el ítem
1	3.86	1.54	.505	.6863	.500	.7057	.513	.7098
2	4.03	1.43	.522	.6815	.519	.7009	.509	.7103
3	5.44	1.49	.507	.6906	.509	.7054	.511	.7121
4	5.02	1.40	.525	.6867	.525	.7026	.535	.7075
5	5.02	1.55	.457	.6956	.460	.7086	.451	.7187
6	4.36	1.47	.557	.6765	.551	.6972	.556	.7024
7	5.39	1.44	.452	.6901	.449	.7078	.440	.7172
8	4.92	1.38	.528	.6868	.530	.7008	.523	.7090
9	4.95	1.72	<u>.298</u>	.7126	<u>.298</u>	.7293	-	-
10	2.12	1.18	<u>.050</u>	.7234	-	-	-	-
11	4.95	1.24	.547	.6866	.549	.7017	.547	.7089
12	5.08	1.16	.595	.6857	.598	.6992	.610	.7031
13	5.30	1.27	.521	.6949	.526	.7075	.541	.7116
Alfa de Cronbach			.7093		.7234		.7293	

b) Comparación de las puntuaciones en GCS (Estudio 3) y en la ESI (Estudio 4)

Se realizó un ANOVA en el que se introdujeron la escala (GCS vs. ESI) y el grupo de pertenencia (policías noveles vs. policías veteranos) como variables independientes, y las puntuaciones obtenidas en estas escalas como variable dependiente⁵⁰.

Los resultados de esta prueba indicaron un efecto principal de la escala, $F_{(1,321)} = 276.94$, $p < .001$, $\eta^2 = .463$. Tal como se observa en la Figura 9, la puntuación media alcanzada en la escala de GCS ($M = 36.35$, $DE = 10.58$) fue significativamente inferior a la puntuación media alcanzada en la ESI ($M = 53.37$, $DE = 8.01$). Este es un resultado congruente con la idea de que en situaciones que se interpretan como policialmente relevantes, como los interrogatorios, es donde los profesionales muestran mayor desconfianza.

⁵⁰ Nótese que tanto la escala de GCS como la ESI tenían el mismo número de ítems (11) y se respondían en escalas idénticas, por lo que podían compararse directamente, sin necesidad de promediar las puntuaciones.

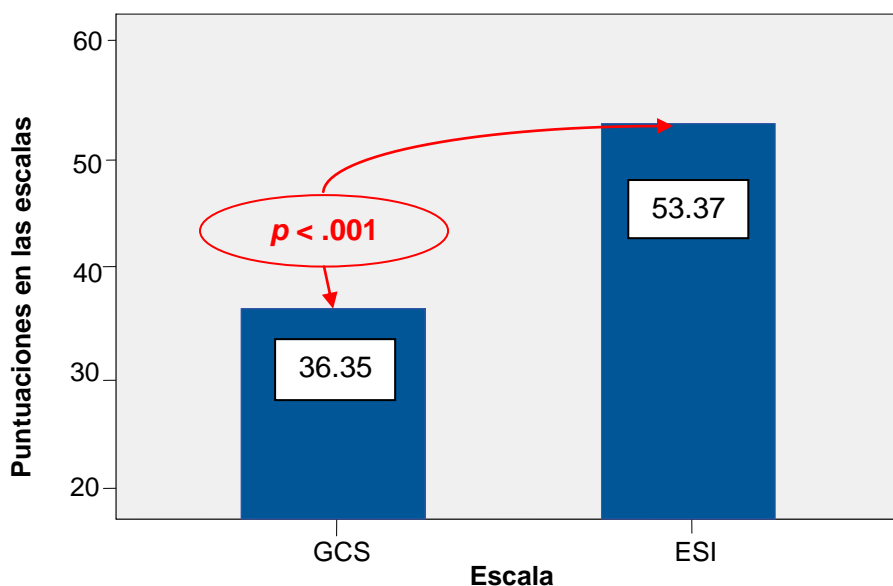


Figura 9. Puntuaciones Medias Obtenidas en las Escalas de GCS (Estudio 3) y Sospecha en Interrogatorios (Estudio 4).

Se observó, además, un efecto significativo del grupo de pertenencia, $F_{(1,321)} = 4.08$, $p = .044$, $\eta^2 = .013$. Como se observa en la Figura 10, la puntuación media total (en ambas escalas tomadas conjuntamente) de los policías noveles ($M = 43.99$, $DE = 13.36$) fue inferior a la mostrada por los policías con experiencia ($M = 47.32$, $DE = 11.62$). Sin embargo, este efecto se vio matizado por una interacción significativa entre el Grupo de pertenencia y la Escala, $F_{(1,321)} = 7.60$, $p = .006$, $\eta^2 = .023$ (ver Figura 11). Esto ofrece apoyo a nuestra hipótesis.

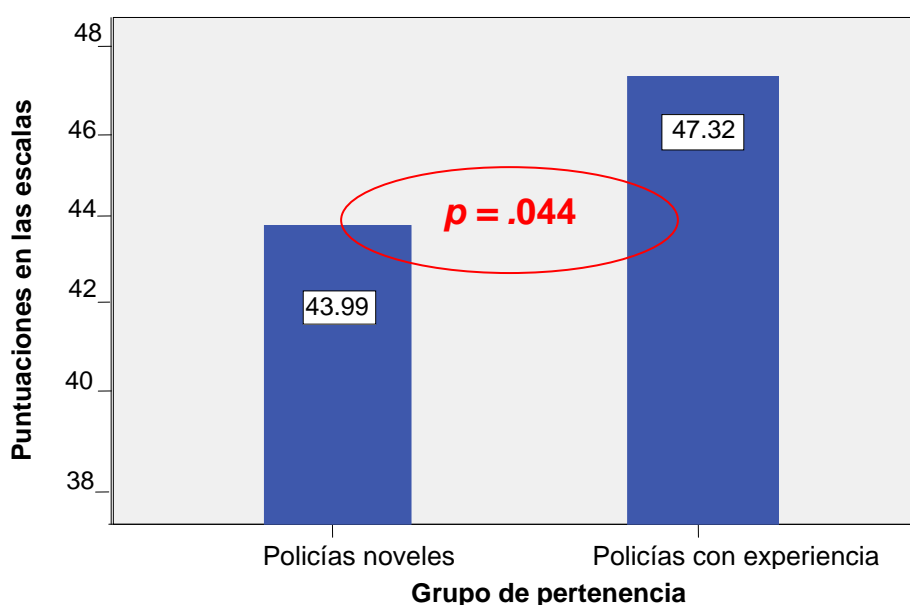


Figura 10. Puntuaciones en GCS y Sospecha en Interrogatorios.

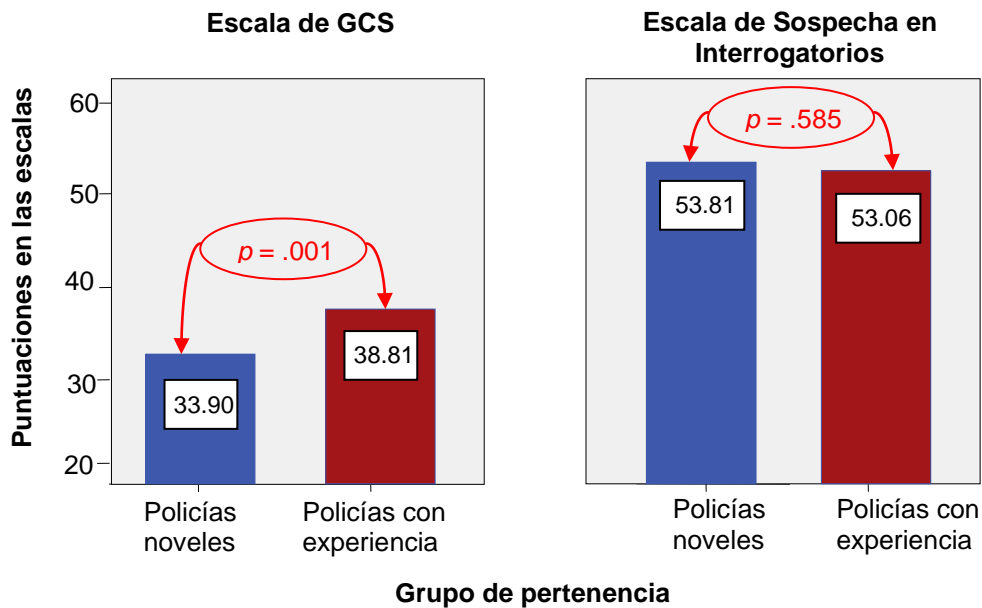


Figura 11. Puntuaciones en las Escalas de GCS y Sospecha en Interrogatorios.

Conclusiones

Los resultados de este estudio son consistentes con la idea que se ha propuesto de que, cuando los policías noveles interpretan una situación como relevante para la función policial (por ejemplo, los interrogatorios), hacen tantos juicios de mentira como los policías veteranos. La construcción que los policías noveles hacen de este tipo de situaciones (como un contexto en el que se debe sospechar fuertemente de la veracidad de las declaraciones) podría explicar por qué en la tarea de evaluación de la credibilidad propuesta en el Estudio 3, éstos mostraron unos porcentajes de juicios de mentira similares a los de los policías con experiencia, porcentajes que fueron superiores a los de los estudiantes. Éstos probablemente construyeron la situación de forma diferente, como se deduce de que, al contrario que los policías, mostraran un sesgo hacia la verdad.

CONCLUSIONES DE LA INVESTIGACIÓN SOBRE LA EXPERIENCIA COMO POLICÍA

En esta primera parte del trabajo empírico realizado, se ha tratado de profundizar en una de las dos variables que Meissner y Kassin (2002) sitúan en la base del efecto del sesgo del investigador: la experiencia como profesional.

Para explicar el efecto de la experiencia sobre los juicios de veracidad de los policías se propuso inicialmente una explicación basada en el desarrollo progresivo de una GCS por parte de los policías (ver Anexo F). En concreto, se hipotetizaba que a medida que los policías van adquiriendo más experiencia profesional, al participar repetidas veces en interacciones de tipo competitivo caracterizadas por una alta sospecha estado, desarrollarán paralelamente una sospecha comunicacional generalizada. Esta GCS determinará un sesgo hacia la mentira, de forma que ante una situación en la que deben juzgar la veracidad de una declaración, los policías con más experiencia (y por lo tanto con mayores niveles de GCS) emitirán más juicios de mentira que aquellos con menos años de experiencia (que no habrán desarrollado aún altos niveles de GCS). De esta forma, la experiencia profesional, tal como Meissner y Kassin (2002) planteaban, en vez de mejorar la precisión en los juicios, únicamente produciría un aumento de juicios de mentira.

A lo largo de los tres primeros estudios que se han presentado, se ha tratado de contrastar empíricamente la explicación anterior. Las principales conclusiones alcanzadas al respecto han sido las siguientes: (a) tal como se esperaba, los policías con más experiencia mostraron mayores niveles de GCS que sus compañeros noveles y que los no policías, lo que nos lleva a concluir que, en general, la experiencia policial aumenta los niveles de GCS; (b) sin embargo, no se halló ninguna correlación significativa entre GCS y el porcentaje de juicios de mentira; y (c) ante una tarea en la que se debía juzgar la veracidad de una serie de declaraciones, tanto los policías noveles como los veteranos mostraron un sesgo hacia la mentira, mientras que un grupo de estudiantes mostró un sesgo hacia la verdad. La ausencia de diferencias entre policías noveles y expertos en términos de juicios de mentira, así como la ausencia de relación entre puntuaciones en GCS y juicios de mentira, ponen en entredicho que la GCS sea un buen predictor de los juicios de mentira.

¿Por qué los policías noveles hicieron tantos juicios de mentira como los veteranos? Esta es la pregunta que nos hicimos a la vista de los resultados encontrados. Presentamos entonces una posible explicación a este interrogante partiendo no de la GCS, sino de cómo los policías noveles pueden interpretar que deben comportarse en una situación que han construido como

relevante para su trabajo policial. En concreto se planteaba que en determinadas situaciones propias del trabajo policial, tales como interrogatorios o tareas que implican juzgar si la declaración de alguien es verdad o mentira, los policías interpretarán que deben adoptar un rol profesional coherente con la idea o estereotipo de que, como policías, su labor consiste en sospechar y desconfiar del otro, pues su objetivo es resolver casos, es decir, encontrar culpables (mentirosos en este caso). En consonancia con este nuevo planteamiento, los resultados del Estudio 4 muestran que: (a) en situaciones propias del trabajo policial en que se debe juzgar si alguien miente o dice la verdad, los policías entienden que deben mostrar una elevada sospecha ante las declaraciones que van a escuchar; y (b) esto sucede independientemente de los años de experiencia que lleven en el Cuerpo. Una comparación de los datos de los Estudios 3 y 4 mostró que la sospecha específica de una situación de interrogatorio es mayor que la GCS que los policías puedan haber desarrollado por sus años de experiencia.

De esta forma, si los policías noveles percibieron la tarea de evaluación de la credibilidad como profesionalmente relevante, es posible que fuera la forma en que construyen este tipo de situaciones y cómo debe ser su rol (desconfiar y cuestionar fuertemente la declaración de la persona interrogada) la que guiara sus juicios. Por su parte los estudiantes, que probablemente no asumieron el mismo rol profesional que los policías, no tuvieron porqué identificar la tarea de evaluación de la credibilidad del Experimento 3 como una ocasión en la que debían desconfiar, sospechar y buscar mentiras. De hecho, los datos apuntan más bien en dirección contraria, dado el sesgo hacia la verdad que mostraron en sus juicios.

Este planteamiento, basado en la construcción de la situación y el rol profesional del policía, se muestra así como una explicación plausible de porqué los policías noveles emitieron tantos juicios de mentira como sus compañeros veteranos en la tarea de evaluación de la credibilidad del Estudio 3, a pesar de las diferencias en GCS de ambos grupos.

El efecto del entrenamiento en la detección del engaño sobre los juicios de veracidad de los policías **7**

El segundo factor que Meissner y Kassin (2002) sitúan en la base del efecto del sesgo del investigador es el entrenamiento para la detección del engaño. Estos autores han planteado que los programas de entrenamiento, lejos de mejorar las tasas de precisión en tareas de evaluación de la credibilidad, aumentan las probabilidades de juzgar una declaración como falsa.

A continuación nos centraremos en el entrenamiento desde una doble perspectiva: en primer lugar, se expondrá cómo es el entrenamiento que los policías reciben en sus centros de formación y de trabajo; en segundo lugar, se examinarán los resultados de distintos estudios de laboratorio en los que se ha ofrecido algún tipo de entrenamiento diseñado a partir de los resultados de la investigación. Se ha decidido establecer esta distinción porque, a menudo, los programas de entrenamiento que se han utilizado en la investigación de laboratorio difieren en gran medida de los que los policías reciben y de los que recomiendan los manuales a los que suelen acceder. Finalmente, se presentará la investigación que hemos realizado sobre el entrenamiento.

EL ENTRENAMIENTO QUE RECIBEN LOS POLICÍAS EN LOS CENTROS DE FORMACIÓN Y DURANTE LOS PRIMEROS AÑOS COMO PROFESIONALES

El objetivo de este apartado es examinar en primer lugar en qué medida los policías reciben entrenamiento en la detección del engaño durante su periodo de formación y en sus primeros años como profesionales. Se expondrán seguidamente los parámetros de la *Técnica Reid*, dado que se trata de uno de los protocolos más conocidos al que recurren los profesionales de distintos países a la hora de buscar orientación sobre cómo realizar interrogatorios y detectar el engaño. Se observarán también las directrices incluidas en el manual de formación de los policías españoles.

LA FORMACIÓN EN DETECCIÓN DEL ENGAÑO ES ESCASA

Dada la importancia que supone para los policías saber distinguir cuando alguien está mintiendo y cuando diciendo la verdad, resultan sorprendentes las críticas de varios investigadores del campo de la detección del engaño que afirman que los policías reciben poco entrenamiento sobre evaluación de la credibilidad o sobre la obtención de información a través de un interrogatorio.

Algunos datos sobre la formación en el Reino Unido

Ray Bull indica en distintos trabajos (por ejemplo Bull, 1989; véase también Bull, 1999) que los policías ingleses reciben poco entrenamiento en la detección del engaño. En el segundo capítulo de su célebre manual sobre interrogatorios policiales y confesiones, Gudjonsson (2003) revisa los estudios que han examinado cómo se realizan los interrogatorios policiales en el Reino Unido. La primera idea que plantea el autor es que hasta principios de 1990 no se ofrecía ningún tipo de entrenamiento a los policías británicos. A día de hoy, algunos de los investigadores más importantes del Reino Unido siguen haciendo referencia en sus publicaciones al hecho de que en este país los policías no reciben ningún entrenamiento en detección del engaño (Vrij et al., 2008).

Algunos datos sobre la formación en Estados Unidos

Desde los Estados Unidos, Frank y Feeley (2003) también denuncian el escaso entrenamiento que reciben los policías norteamericanos. El primer autor afirma haber trabajado con alrededor de 1000 profesionales de las fuerzas de la ley, de los que tan solo un 10% había recibido algún tipo de entrenamiento específico. De este 10% supuestamente entrenado, a muchos simplemente se les había presentado una lista muy simplificada de indicios conductuales cuya utilidad real, según indica el autor, no ha sido corroborada por la evidencia científica.

Frank y Feeley (2003) plantean cuatro posibles razones por las que los agentes de la ley norteamericanos reciben poco entrenamiento en detección del engaño. En primer lugar, dado que la literatura científica sugiere que no hay

indicios perfectos de la mentira, los policías podrían negarse a recibir entrenamiento. En segundo lugar, y en relación con el punto anterior, los indicios útiles solamente tienen un valor probabilístico, y quizá los policías no estén dispuestos a utilizarlos sabiendo que pueden hacer que alguien que miente sea juzgado como veraz. Además, Frank y Feeley hacen referencia a un cierto rechazo de los profesionales de las fuerzas de la ley a los académicos que presentan los resultados de la investigación en términos probabilísticos. Quizá los policías entiendan que la ciencia es una herramienta útil solamente si les ofrece métodos infalibles de detección, lo cual a día de hoy es imposible. Finalmente, según los autores, muchos policías aseguran que la tasa base de mentiras con las que se enfrentan en sus interrogatorios es superior al 75%. Si este dato fuera real (cosa de la que no hay ninguna evidencia científica), un policía que juzgase a toda persona que interrogase como mentirosa tendría una tasa de precisión también superior al 75%.

En un trabajo posterior, Colwell, Miller, Lyons y Miller (2006) examinaron, vía cuestionario, el tipo y cantidad de entrenamiento que reciben policías norteamericanos de distintas agencias de Texas, tanto al principio de su carrera como más adelante. Respecto al entrenamiento inicial, los datos mostraron que solo el 56% había recibido algún tipo de entrenamiento formal⁵¹, bien durante su periodo de formación en la academia (30%), bien asistiendo a algún tipo de seminario (26%). En la mayoría de los casos, el entrenamiento recibido consistía en talleres de lectura de dos días de duración (ocho horas al día) en los que solamente se proporcionaba a los policías información escrita sobre cómo interrogar a los sospechosos. Otro 34% afirmó haber sido aleccionado en su lugar de trabajo, normalmente por otros policías con más años de experiencia en el Cuerpo. El 10% restante indicó no haber recibido ningún tipo de entrenamiento, o haber tenido contactos con facultades, universidades, o haber pasado por experiencias personales que les habían servido de ayuda.

⁵¹ Al hablar de entrenamiento formal, los autores se refieren a la instrucción específica en la academia o a seminarios y cursos destinados específicamente al entrenamiento en detección del engaño.

En lo referente al entrenamiento posterior al periodo de formación inicial, sólo una minoría (33%) de los policías encuestados informó haberlo recibido (véase Colwell, Miller, Lyons, et al., 2006, para una información más detallada).

En definitiva, según este estudio casi la mitad de los policías encuestados no había recibido entrenamiento formal. Además, en aquellos casos en los que se había ofrecido entrenamiento, éste era claramente insuficiente en cuanto al formato y el tiempo dedicado.

Algunos datos sobre la formación en España

Por lo que se refiere a los policías españoles, no tenemos constancia de ningún estudio similar al de Colwell, Miller, Lyons, et al. (2006) realizado en nuestro país. Según el Plan de Estudios elaborado el año 2000 por la División de Formación y Perfeccionamiento de la División General de la Policía, vigente hasta el 2006, la formación de los alumnos de la Escala Básica, se organizaba en distintos módulos. Cada módulo estaba compuesto por distintas unidades didácticas, y cada unidad didáctica costaba de distintos temas. Considerando la relación entre el número de horas de formación y el número de temas, a cada tema le corresponderían aproximadamente cinco horas de formación⁵². Del total de temas incluidos en el Plan de Estudios, solamente uno se refería explícitamente al testimonio, incluyéndose un apartado sobre la detección del engaño y otro sobre el interrogatorio policial. Es decir, la formación de los policías de la Escala Básica incluía, teóricamente, alrededor de cinco horas que se repartían entre estos dos apartados.

La formación de los policías de la Escala Ejecutiva se organizaba por asignaturas, cada una de las cuales constaba de distintos créditos⁵³. Solamente la asignatura de “Investigación e Información” presentaba un tema en el que, explícitamente, se incluía un apartado dedicado a la entrevista y el interrogatorio policial, lo que suponía, en función de los créditos de la asignatura y el número de temas, alrededor de tres horas y media de formación.

⁵² Este es un cálculo meramente aproximativo, ya que nos consta que en la mayor parte de las unidades didácticas algunos temas se alargaban a costa de reducir el tiempo dedicado a otros.

⁵³ Un crédito equivalía 10 horas de formación.

Es necesario reiterar que los datos que se acaban de presentar son meramente aproximativos, ya que no se ha podido disponer directamente de los temarios detallados de las asignaturas, ni de los contenidos específicos de cada tema.

Adicionalmente, tanto los alumnos de la Escala Básica como de la Escala Ejecutiva pueden participar voluntariamente en distintos cursos y seminarios a lo largo del año académico. No hemos podido acceder a los contenidos de los mismos, aunque sí nos consta que en algún caso se trató la obtención de información en los interrogatorios. Finalmente, hay que tener en cuenta que durante el periodo comprendido entre los años 2000 y 2006 (año en que se presentó un nuevo plan de estudios), aunque la organización de las asignaturas y módulos fuera similar, es posible que muchas materias modificasen sus contenidos, introduciendo, modificando o eliminando algunos temas. Por ejemplo, Carrancho (2002) incluyó en el manual de la asignatura “Psicología Social II” del segundo curso de la Escala Ejecutiva dos nuevos temas, uno dedicado a la entrevista policial y otro a la psicología del testimonio, en el que se incluía un apartado dedicado a la detección del engaño.

En el año 2006, la División de Formación y Perfeccionamiento presentó un nuevo Plan de Estudios que cambió la formación de los policías de la Escala Ejecutiva, tanto respecto a los contenidos como a la forma de organizarlos. Algunas asignaturas desaparecieron, se crearon otras nuevas y se potenciaba la participación en cursos y seminarios, tanto durante el proceso de formación en el Centro de Ávila, como una vez finalizado este periodo.

Con el nuevo Plan de Estudios, se ha tratado de dar un giro hacia la potenciación de los conocimientos aportados desde la ciencia, especialmente desde la Psicología. En este sentido, los policías de la Escala Ejecutiva disponen, por ejemplo, de la asignatura “Base Científica de la Investigación y Diagnóstico de Criminalidad”, en la que se les instruye sobre cómo aplicar el método de investigación utilizado en ciencia a la investigación policial, se les enseña a elaborar un cuestionario, a planificar una entrevista o a aplicar la observación científica a la actividad policial, así como a manejar la estadística para fines de investigación. Adicionalmente, en la asignatura “Interacción Social

(Habilidades Sociales)”, se ofrece formación sobre procesos psicosociales básicos (por ejemplo, aprendizaje y emoción), personalidad y diferencias individuales, procesos de influencia social (categorización, atribución de causalidad o profecías autocumplidas) o entrenamiento en habilidades sociales (técnicas comunicativas asertivas).

A pesar de todo, la formación en detección del engaño sigue siendo muy escasa, ya que el Plan de Estudios del 2006 ha suprimido las asignaturas del anterior Plan en las que se ofrecía información sobre el tema (“Investigación e Información” y “Psicología Social II”). No obstante, la detección del engaño se trata en dos asignaturas. En una de ellas, “Gestión de Recursos Humanos en el C.N.P.”, se incluye un tema específicamente dedicado a la entrevista policial y el interrogatorio, y otro al testimonio. En ellos se ofrece información sobre la Entrevista Cognitiva, y se incluye un apartado referido a la detección del engaño. Adicionalmente, en la ya mencionada asignatura “Base Científica de la Investigación y Diagnóstico de Criminalidad” se tratan también temas referidos a la obtención de información mediante entrevistas e interrogatorios. No obstante, al igual que se ha indicado respecto al Plan de Estudios del año 2000, la información que aquí se presenta sobre el actual Plan es meramente aproximativa, ya que no hemos podido disponer ni de los temarios detallados de las asignaturas, ni de los contenidos específicos de cada tema.

Conclusiones generales

En definitiva, se puede concluir que, tal como indican Colwell, Miller, Lyons, et al. (2006), la formación que reciben los policías norteamericanos sobre detección del engaño durante su estancia en los Centros y Academias de Formación es, en general, escasa. Un problema, quizá derivado de esta situación, es el que Bull (1999) plantea respecto a los policías británicos: ante la falta de formación, dichos policías pueden haberse decantado por la adopción de ciertas técnicas de detección de mentiras, presentadas en manuales, publicitadas como “altamente eficaces” pero cuya validez no ha sido científicamente probada. O, peor aún, por técnicas que la evidencia científica ha mostrado ineficaces. Respecto a los policías españoles, aunque su formación, a nivel global, parece ir avanzando progresivamente hacia a una

perspectiva más científica, la instrucción específica sobre entrevistas policiales, interrogatorios y, especialmente, detección del engaño es aún incompleta, tanto respecto a las horas dedicadas a tales temas como a la profundidad de los contenidos.

EL MERCADO DE LOS MANUALES: LA *TÉCNICA REID*

Como se planteó al final del epígrafe anterior, la escasez de entrenamiento en detección del engaño durante el periodo de formación de los policías, así como la falta de una técnica o protocolo estandarizado, han sido dos importantes razones para que muchos profesionales puedan haber optado como alternativa por los distintos manuales que se ofertan en el mercado (véase Bull, 1989, 1999). En este sentido, la técnica a la que más se ha recurrido en los Estados Unidos (y que, según Bull, 1999, también se ha difundido ampliamente en el Reino Unido) es la *Técnica Reid*, que se presenta en el manual de Inbau, Reid, Buckley y Jayne (2001), que ya cuenta con varias ediciones en el mercado. Dada la repercusión e influencia de esta técnica sobre la práctica policial de profesionales de múltiples países, es obligado examinarla con cierto detalle.

Este manual incluye, en primer lugar, una serie de líneas guía sobre cómo realizar una entrevista policial (la *Behavior Analysis Interview* o *BAI*). A partir de esta entrevista se toma una decisión sobre la sinceridad de la persona entrevistada en base a la observación de distintas conductas verbales y no verbales. Si el entrevistador concluye que el entrevistado miente, entonces se decide someterlo a un interrogatorio. Dicho interrogatorio, que también se describe en el manual de Inbau et al. (2001), se estructura como una sucesión de nueve pasos.

Este procedimiento no sólo se describe en el manual de Inbau et al. (2001), sino que, además, constituye el núcleo de los programas de entrenamiento en entrevistas e interrogatorios de la empresa *John E. Reid & Associates*. Dicha empresa afirma haber entrenado a más de 500.000 profesionales de distintos ámbitos legales (véase http://www.reid.com/training_programs/interview_overview.html). Policías de los Estados Unidos, Canadá, Méjico, Alemania,

Bélgica, Japón, Corea del Sur, Singapur, Arabia Saudí y Emiratos Árabes han sido entrenados con esta técnica (Blair y Kooi, 2004).

La *Behavior Analysis Interview (BAI)* como instrumento para evaluar la credibilidad

La *BAI* incluye una serie de *preguntas de investigación* (específicas para cada caso) y *preguntas generadoras de conducta*. Como indican los propios autores, “el núcleo de la *Behavior Analysis Interview* ... es el planteamiento de *preguntas generadoras de conducta*, que están dirigidas específicamente a generar respuestas conductuales” (Inbau et al., 2001, p. 173). Se trata de 15 preguntas estandarizadas que se adaptan a cada caso específico. Los autores han etiquetado cada una de ellas dependiendo de su función en la entrevista: *Objetivo, historia / usted, conocimiento, sospecha, garantía, credibilidad, oportunidad, actitud, pensamiento, motivo, castigo, segunda oportunidad, objeción, resultados y contarlo a los seres queridos* (véase Inbau et al., 2001, para una explicación del propósito de cada pregunta). La premisa de la que se parte es que las 15 preguntas elicitarán respuestas distintas en quienes mientan y en quienes digan la verdad. En el Anexo L se presenta un ejemplo de cada una de tales preguntas adaptadas a un caso específico, y las respuestas y reacciones que se predicen en función de si el entrevistado miente o dice la verdad.

Indicios no verbales del engaño según la *Técnica Reid*

Además de las predicciones sobre la conducta verbal y no verbal de los culpables que se establecen específicamente desde la *BAI* (Anexo L), Blair y Kooi (2004) indican también que existen, a nivel no verbal, una serie de áreas o zonas de evaluación que, según la *Técnica Reid* el policía debe observar mientras realiza la entrevista, y que, según esta técnica mostrarán muy probablemente si el sospechoso miente o dice la verdad:

a) Postura. De acuerdo con la *Técnica Reid*, quienes digan la verdad permanecerán más erguidos, derechos, alineados frontalmente con el entrevistador y sentados más cómodamente, mientras que quienes mientan tenderán más a recostarse sobre su asiento y mostrarán menos frontalidad con

el entrevistador. Además, adoptarán posturas a las que Blair y Kooi (2004) se refieren como barreras, por ejemplo cruzar los brazos o las piernas de forma tensa.

b) Las manos y los brazos. Se espera que quienes digan la verdad hagan más movimientos de apoyo a su argumentación verbal (ilustradores), y quienes mientan muestren más adaptadores, tales como arreglarse el pelo o la ropa, frotarse las manos, o golpear rítmicamente los dedos.

c) Los pies. Se predice que la presencia de movimientos rítmicos y repetitivos de los pies puede indicar engaño. Adicionalmente, se plantea que los movimientos de los pies y las piernas están implicados en los cambios de postura, y se considera un signo de engaño a los cambios de postura que ocurren consistentemente cada vez que el investigador formula distintas preguntas críticas sobre un tema concreto a lo largo de la entrevista.

d) El contacto ocular. Se sostiene que los entrevistados que digan la verdad mantendrán el contacto ocular con el entrevistador de forma normal (de acuerdo con la *Técnica Reid*, “lo normal” es entre un 30% y un 60% del tiempo), mientras que quienes mientan tenderán bien a evitar el contacto ocular, o bien a mantener la mirada excesivamente fija en el entrevistador en un intento de compensar el deseo de mirar a otro lado.

e) La conducta paralingüística. Se plantea que hay que observar la duración de la respuesta (se predice que las personas que mienten emiten respuestas más cortas), la latencia de respuesta (se predice que quienes mienten mostrarán una mayor latencia de respuesta), la tasa de habla (se indica que las respuestas de quienes dicen la verdad incluirán más palabras por minuto que las respuestas de quienes mienten), y el tono de voz (se predice que durante respuestas que impliquen una carga emocional, quienes dicen la verdad mostrarán un incremento en el tono de voz).

El interrogatorio

De esta forma, la *Técnica Reid* plantea que el entrevistador debe observar las conductas verbales, no verbales y paralingüísticas del sospechoso. En base

a las respuestas del interrogado a la *BAI*, y a la observación de su conducta durante dicha entrevista, el policía debe tomar una decisión sobre la veracidad de la declaración. Si finalmente se considera que el entrevistado ha mentido y está implicado de alguna forma en el delito, entonces se procederá a interrogar a esa persona. El manual de Inbau et al. (2001) propone para ello un interrogatorio muy coercitivo que consta de nueve pasos, y que por sus características, como por ejemplo la presión a la que se somete al interrogado durante todo el proceso, puede generar falsas confesiones.

¿Funciona realmente la *Técnica Reid*?

La pregunta natural que cabe formularse es si realmente éste método funciona. Horvath, Jayne y Buckley (1994) examinaron la utilidad de la *BAI* en un estudio de campo en el que cuatro personas evaluaron 60 *BAI*s grabadas en vídeo (30 verdades y 30 mentiras) que habían sido administradas por cinco entrevistadores. Sin considerar los casos inconcluyentes (aquellos en los que no se tomó ninguna decisión respecto a la veracidad) la precisión fue del 91% para las verdades y del 80% para las mentiras. No obstante, cuando se incluían los casos inconcluyentes, las tasas de precisión se situaron en el 78% para las verdades y el 66% para las mentiras. Aunque este estudio ofrece apoyo empírico a la *BAI*, ha sido criticado por sus limitaciones metodológicas (véase, por ejemplo, Vrij, Mann y Fisher, 2006). La más importante se refiere al modo en que se estableció la veracidad real de las declaraciones. Si no podemos estar seguros de si las declaraciones que los autores consideraron verdaderas eran realmente verdaderas, ni de si las declaraciones consideradas falsas eran realmente falsas, entonces no podemos conocer con exactitud los niveles reales de precisión.

Kassin y Fong (1999) realizaron un estudio con el propósito de comprobar en qué medida el entrenamiento en el empleo de los indicios verbales y no verbales del engaño de la *Técnica Reid* mejoraba la tasa de precisión al evaluar la credibilidad. En concreto, se tomó un grupo de 16 estudiantes. A ocho se les instruyó para que cometieran uno de cuatro posibles falsos delitos (véase Kassin y Fong, 1999, para una explicación detallada), y a los otros ocho se les pidió que realizaran tareas similares, pero no delictivas. Posteriormente,

se informó a los 16 estudiantes que serían interrogados sobre la comisión de un delito y que debían negar su implicación. El contexto en el que se realizó el interrogatorio simulaba las condiciones recomendadas por el manual de Inbau et al. (1991), aunque, obviamente, no se ejerció sobre los participantes la intensa presión que se recomienda en el manual. Las declaraciones de los 16 interrogados se grabaron en vídeo y se mostraron a un grupo de 40 estudiantes de psicología (véase Kassin y Fong, 1999). A la mitad se le entrenó para que juzgase las declaraciones utilizando los indicios verbales y no verbales de la *Técnica Reid*. A la otra mitad no se le dio ningún tipo de entrenamiento.

Se encontró que los participantes no entrenados alcanzaron tasas de precisión superiores a las de los entrenados. Sin embargo, estos últimos mostraron más confianza en sus juicios. Los autores concluyen que la *Técnica Reid* tiene efectos contraproducentes, debido a que los indicios verbales y no verbales de este sistema carecen de valor diagnóstico.

En este sentido, Blair y Kooi (2004) compararon las predicciones de la *Técnica Reid* sobre los indicios de engaño con los resultados del meta-análisis de DePaulo, Lindsay, et al. (2003). Las escasas coincidencias llevaron a los autores a la siguiente conclusión:

En general, los resultados sugieren que el modelo de *Reid* en relación a la conducta no verbal es demasiado simplista, y en algunos casos simplemente incorrecto. Uno podría esperar encontrar que las personas entrenadas con la técnica *Reid* podrían exhibir una menor precisión en la detección del engaño (Blair y Kooi, 2004, p. 82).

En otra interesante investigación, Vrij, Mann y Fisher (2006) contaron con 40 estudiantes, de los que 20 mintieron y 20 dijeron la verdad mientras eran entrevistados con la *BAI* sobre una serie de acontecimientos ocurridos durante una supuesta sesión experimental (véase Vrij, Mann y Fisher, 2006, para más detalles). Las 40 entrevistas se grabaron en audio, vídeo y se transcribieron en papel.

El propósito de los autores era comparar las predicciones que Inbau et al. (2001) establecen respecto a la conducta verbal y no verbal de quienes mienten en la *BAI* con las predicciones derivadas de la *Impression Management Hypothesis* (Hipótesis del Manejo de la Impresión⁵⁴, en adelante HMI) de DePaulo (1992). Desde esta teoría, y de forma radicalmente opuesta a los planteamientos de Inbau et al. (2001), se hipotetiza, a nivel general, que (a) las respuestas de los mentirosos parecerán más útiles y amables que las de aquellos que dicen la verdad y (b) la conducta no verbal de los mentirosos parecerá en general más relajada que la de quienes dicen la verdad (véase el Anexo M para una comparación más detallada de las predicciones hechas desde la HMI y desde la *BAI* sobre la conducta de quienes mienten y dicen la verdad).

Para comparar qué predicciones eran más correctas, los autores desarrollaron un cuestionario que medía las diferencias en la conducta verbal (a partir de las transcripciones de las entrevistas) y no verbal (a partir de las grabaciones en formato vídeo) de los entrevistados. Dicho cuestionario se construyó de tal forma que si las predicciones de la *BAI* eran correctas, entonces se obtendría una puntuación mayor para las personas que habían dicho la verdad durante la entrevista que para quienes habían mentido (véase Vrij, Mann y Fisher, 2006). Una vez construido el cuestionario, dos codificadores lo usaron para codificar las transcripciones de las entrevistas.

Los resultados mostraron que las puntuaciones globales en el cuestionario para quienes mentían fueron más altas que las puntuaciones para quienes decían la verdad. Es decir, las conductas mostradas por las personas que mentían y decían la verdad durante la entrevista se ajustaban a las predicciones establecidas desde la HMI, e iban en dirección radicalmente opuesta a las predicciones establecidas desde la *BAI*. Al examinar las preguntas de la *BAI* individualmente, Vrij, Mann y Fisher (2006) encontraron

⁵⁴ Al introducir el estudio de DePaulo, Lindsay, et al. (2003) en el apartado referido a los indicios no verbales del engaño, ya nos referimos a la teoría de la auto-presentación o del manejo de la impresión. Recordamos ahora, no obstante, que desde esta teoría se hipotetiza que las personas adecuamos nuestra conducta a nuestros intereses en cada situación. En este sentido, una persona que miente pero que desea ser creída tratará de controlar su conducta para mostrarse a los demás tal y como cree que se comporta alguien sincero. Para una exposición más detallada de esta teoría, el lector interesado puede acudir a DePaulo (1992).

que las diferencias conductuales reales iban en sentido opuesto a lo hipotetizado por Inbau et al. (2001) en cuatro de ellas; en el resto, las diferencias no eran significativas. Los autores concluyen con las siguientes palabras:

La *BAI* es una técnica de entrevista desarrollada por Inbau et al. (2001) y diseñada para evocar diferentes respuestas verbales y no verbales en personas mentirosas y veraces. Efectivamente, la entrevista *BAI* generó diferencias entre mentirosos y veraces, pero tales diferencias fueron en sentido opuesto a las predicciones de Inbau et al. Ellos esperaban que los mentirosos colaboraran menos en las investigaciones y exhibieran más conductas nerviosas. Sin embargo, los mentirosos fueron más colaborativos y mostraron menos conductas nerviosas (Vrij, Mann y Fisher, 2006, p. 342).

Finalmente, el estudio de Vrij et al. (2007), al que ya se ha hecho referencia anteriormente (ver el apartado dedicado a la precisión de los policías como detectores de mentiras), encontró que la *BAI* generaba menos indicios útiles para la discriminación de verdades y mentiras que una entrevista facilitadora, caracterizada por el empleo de preguntas abiertas. Además, se observó que juzgar la veracidad de las declaraciones obtenidas mediante una *BAI* no implicaba un aumento de la precisión, pero sí un aumento de la confianza en los propios juicios.

En definitiva, los resultados de los estudios expuestos apuntan a que el entrenamiento en la *Técnica Reid* no solamente no mejora la precisión en la evaluación de la credibilidad, sino que la empeora (Kassin y Fong, 1999). Esto pudiera deberse a que los indicios recomendados para juzgar la veracidad de una declaración no son adecuados (Blair y Kooi, 2004). De hecho, las predicciones que se hacen sobre cómo será la conducta de las personas que mienten parecen ir en dirección contraria a la forma en que realmente se comportan los individuos al mentir (Vrij, Mann y Fisher, 2006). Además la *BAI* no hace que los entrevistados muestren tantos indicios conductuales útiles para la distinción de verdades y mentiras como otros formatos de entrevista, como la facilitadora (Vrij et al., 2007).

Así pues, si los policías deben decidir si alguien debe ser interrogado a partir de un juicio de veracidad basado en la *BAI*, es probable que se juzgue a muchas personas veraces como mentirosas, y que se las someta a un interrogatorio. De esta forma, hay personas inocentes serán sometidas a interrogatorios en los que se les presupone culpabilidad, se les interrogará con un estilo claramente acusatorio, y se tratará de romper su resistencia para lograr una confesión. Uno de los científicos sociales que mejor y en más reiteradas ocasiones se ha referido a las consecuencias que tiene todo este proceso, y en especial al modo fuertemente inculpatario en que se enfoca el interrogatorio desde el manual de Inbau et al. (2001), es S. M. Kassin, que en numerosas publicaciones ha denunciado lo relativamente fácil que es obtener falsas confesiones utilizando la *Técnica Reid* (Kassin, 2002; Kassin y College, 2005; Kassin et al., 2002; Kassin et al., 2007; Kassin et al., 2005).

La pregunta que daba título a este epígrafe era: ¿funciona realmente la *Técnica Reid*? Se puede contestar a esta cuestión con una respuesta sencilla y preocupante al mismo tiempo: la entrevista *BAI*, quizá una de las más empleadas por los policías para juzgar si un individuo miente o dice la verdad, no es una herramienta fiable, y por lo tanto todo el proceso posterior puede estar viciado.

LOS POLICÍAS ESPAÑOLES

¿Conocen los policías españoles la *Técnica Reid*? Es probable, ya que en las estanterías de la biblioteca del Centro de Formación de la Policía de Ávila se encuentra el manual de Inbau et al. (2001). No obstante, los policías españoles cuentan, para su periodo de formación, con las directrices que Carrancho (2002) plantea en el manual de la asignatura Psicología Social II a la hora de realizar una entrevista policial, un interrogatorio y detectar el engaño⁵⁵. Así, se indica cómo detectar la mentira a partir de los cambios fisiológicos y conductuales observados en el individuo cuyo testimonio debe ser evaluado. Respecto a la aproximación fisiológica, se hace una exposición general del polígrafo, apelando a favor de la GKT. En relación a la aproximación

⁵⁵ Aunque el actual Plan de Estudios ha eliminado la asignatura en la que se presentaban estas técnicas, los alumnos pueden seguir accediendo a esta información en la biblioteca del Centro de Formación.

conductual, se ofrece (de forma muy breve) una relación de conductas que se indican asociadas a la mentira. El problema es que algunas de las que se citan (por ejemplo el contacto visual) no se han mostrado realmente discriminativas.

CONCLUSIÓN SOBRE EL ENTRENAMIENTO QUE RECIBEN LOS POLICÍAS

Si el entrenamiento que reciben los policías en detección del engaño durante su formación o el que se presenta en los manuales a los que pueden recurrir fuera realmente eficaz, la consecuencia inmediata sería que estos profesionales serían unos buenos detectores de verdades y mentiras. Sin embargo, la investigación sugiere que el entrenamiento es insuficiente o inadecuado. Hay dos hallazgos principales que sostienen esta conclusión. Tales hallazgos se refieren a las creencias de los policías sobre los indicadores del engaño y a su precisión al hacer juicios de veracidad.

En primer lugar, distintos estudios han encontrado que las creencias de los policías sobre los indicios conductuales del engaño a menudo no coinciden con los indicadores reales de engaño. Esto indica que los policías o bien no han recibido formación sobre el particular, o bien dicha formación ha sido errónea.

En este sentido, Strömwall, Granhag y Hartwig (2004), a partir de la revisión de diez estudios, elaboraron una lista de las conductas que los policías estiman asociadas a la mentira (Tabla 32).

Tabla 32. *Conductas que los Profesionales Consideran más Indicadoras de Mentiras (Strömwall et al., 2004).*

Aumento de manipuladores
Aumento de movimientos de la cabeza
Aumento de los movimientos de los brazos y las manos
Aumento de movimientos de los pies y las piernas
Aumento de movimientos repetitivos que denotan inquietud
Aumento de la frecuencia de los cambios de postura
Aumento de los movimientos corporales en general
Discurso menos fluido
Evitación de la mirada
Menor consistencia en la argumentación
Menor plausibilidad de las narraciones
Menor cantidad de detalles en las historias

Sin embargo, la investigación sobre indicios reales del engaño ha mostrado que muchas de estas conductas no permiten distinguir verdades de mentiras

(véase el apartado referido a las aproximaciones actuales en la detección de la mentira). La conclusión de los autores es clara y contundente:

Los profesionales tienen básicamente las mismas creencias que personas no profesionales: en general esas creencias son incorrectas, especialmente aquellas referidas a la conducta no verbal. ... El indicador en el que tanto profesionales como no profesionales más confían es la evitación del contacto ocular al mentir, el cual no es un predictor fiable (Strömwall et al., 2004, p. 235).

Más recientemente, Colwell, Miller, Miller y Lyons (2006) examinaron las creencias de 109 policías norteamericanos sobre la medida en que una serie de conductas eran indicativas de engaño. En concreto, se diseñó un cuestionario en el que se presentaba a los participantes un total de 30 indicios conductuales verbales y no verbales y, a través de una escala tipo *Likert* de tres puntos, se les pedía que expresaran sus creencias sobre si cada una de las conductas aumenta, disminuye o no varía al mentir.

Colwell, Miller, Miller, et al. (2006) encontraron que los participantes estimaban que la mayoría de las conductas presentadas aparecían más a menudo al mentir, aún en aquellos casos en los que se trataba de conductas que en realidad no varían (por ejemplo, cambios de postura o pausas durante el discurso), o que disminuyen (gesticulaciones o admisiones de falta de memoria). Una de las creencias erróneas más compartidas (manifestada por el 84% de los policías) era que al mentir disminuye el contacto ocular. Ante estos datos, los autores concluyen que los policías no tienen un conocimiento adecuado de las conductas que realmente se asocian al acto de mentir, y que sus creencias erróneas son similares a las de los no profesionales.

El que los policías tengan creencias erróneas sobre los indicios de engaño constituye un primer indicador de que el entrenamiento que reciben no es adecuado. Otro indicador, posiblemente consecuencia del anterior, lo constituyen las bajas tasas de precisión de los policías al evaluar la credibilidad. Dichas tasas, situadas, como hemos visto ya, cerca del nivel del azar, no son las esperables en estos profesionales.

En definitiva, se puede concluir que (a) los policías no reciben suficiente entrenamiento sobre detección del engaño durante su proceso de formación, o el entrenamiento que reciben es de escasa calidad; (b) esto hace que, en algunos casos, recurran a los manuales que se ofrecen en el mercado, cuya utilidad no ha sido científicamente demostrada; y (c) como consecuencia de lo anterior, los policías muestran creencias erróneas sobre los indicadores del engaño y sus tasas de precisión en tareas de evaluación de la credibilidad no son superiores a las de los no profesionales.

EL EFECTO DEL ENTRENAMIENTO EN LOS ESTUDIOS DE LABORATORIO

Las pobres tasas de precisión (tanto de los profesionales como de los no profesionales) a la hora de evaluar la credibilidad, el hecho de que los policías no reciban un entrenamiento formal adecuado en detección del engaño durante su periodo de instrucción en la academia, y la escasa utilidad de las técnicas ofrecidas en los manuales han sido algunas de las causas que han motivado a muchos investigadores a realizar trabajos para comprobar si un entrenamiento diseñado a partir de los hallazgos de la investigación científica mejora la precisión en la discriminación entre verdades y mentiras.

El objetivo de este apartado es revisar la investigación de laboratorio sobre los efectos del entrenamiento. El primer problema que se va a encontrar alguien que quiera analizar esta cuestión es la diversidad de estrategias y elementos utilizados por los diversos estudios. En este sentido, Vrij (2000) hace una clasificación del formato de entrenamiento que se suele ofrecer en los estudios de laboratorio en tres modalidades. La primera, a la que el autor se refiere como “procedimiento de focalización” (*focusing procedure*), consiste en pedir a los participantes que presten atención a ciertos indicios específicos (los que realmente se asocian a la mentira) y que ignoren otros (que carecen de valor diagnóstico real para detectar el engaño) (por ejemplo, DePaulo, Lassiter y Stone, 1982). La segunda modalidad, a la que Vrij denomina “procedimiento informacional” (*informational procedure*), consiste en proporcionar a los participantes información sobre la verdadera relación entre el acto de mentir y la aparición de ciertas conductas (por ejemplo, Fiedler y Walka, 1993; Vrij,

1994; Vrij y Graham, 1997). Finalmente se propone una tercera modalidad de entrenamiento, consistente simplemente en ofrecer *feedback* a los participantes sobre su ejecución tras emitir cada juicio de veracidad (por ejemplo Fiedler y Walka, 1993; Vrij, 1994).

La clasificación que Vrij (2000) establece de las posibles modalidades de entrenamiento puede, no obstante, hacerse mucho más compleja si se tiene en cuenta que existen notables diferencias dentro de cada modalidad. Por ejemplo las conductas sobre las que dieron información Fiedler y Walka (1993) eran distintas de aquellas sobre las que dieron información Vrij y Graham (1997).

A pesar de esta diversidad, Vrij (2000) revisa el efecto del entrenamiento en 11 estudios. Aunque en la mayoría de los casos el entrenamiento utilizado había sido diferente (tanto respecto al formato como a los indicios conductuales utilizados), el autor concluye que, en general, el entrenamiento produce mejorías muy leves en la precisión de los no profesionales (véase Vrij, 2000). Curiosamente, en los tres estudios realizados con policías, el entrenamiento empeoró las tasas de precisión. Además, en dos de estos estudios (Vrij, 1994; Vrij y Graham, 1997) se observó que los policías entrenados mostraban una confianza en sus juicios ligeramente (aunque no significativamente) superior que la de sus compañeros no entrenados. Así pues, si bien los policías ya tienden a mostrar una gran confianza en sus juicios, ésta podría verse aún más aumentada con el entrenamiento.

Para explicar estos resultados, Vrij plantea que quizá el entrenamiento no mejora la precisión de los policías porque las sesiones de entrenamiento son demasiado complejas y les producen confusión. El autor también especula con la posibilidad de que los policías no se crean la información que reciben durante el entrenamiento.

Posteriormente, Frank y Feeley (2003) presentaron un meta-análisis sobre el efecto del entrenamiento en la detección del engaño. Estos autores basaron su análisis en el tamaño del efecto del entrenamiento (r). Establecieron comparaciones entre 20 grupos que habían recibido alguna forma de entrenamiento y grupos control. Frank y Feeley encontraron que mientras que

en dos casos el tamaño del efecto era superior a .55 (lo que elevó la media a $r = .20$), la mayoría de las veces se obtenía un tamaño del efecto sensiblemente inferior, llegando a ser negativo en varias ocasiones. Con estos datos, los autores señalan que el porcentaje de ganancia en precisión que supone el entrenamiento está en torno al 4%, ($M = 58\%$ para los grupos entrenados vs. 54% para los no entrenados), lo que supone un incremento muy leve. Curiosamente, esta pequeña mejora parece ser consecuencia de un aumento en la precisión al juzgar declaraciones falsas. En ocho de las comparaciones en las que las tasas de precisión se ofrecían por separado para verdades y para mentiras, se observó que la precisión en la detección de verdades entre los participantes entrenados ($M = 56\%$) fue ligeramente inferior que la alcanzada por los participantes no entrenados ($M = 58\%$). Sin embargo, cuando se juzgaban mentiras, los participantes entrenados alcanzaban unas tasas de precisión ($M = 55\%$) superiores a las de los participantes no entrenados ($M = 49\%$). En definitiva, el leve aumento de la precisión global que se ha observado como consecuencia del entrenamiento podría ser consecuencia de un aumento en la precisión en la detección de mentiras, pero no una mejora en la detección de verdades. Quizás, el único efecto del entrenamiento consista en sesgar las respuestas hacia la mentira, lo cual no implica una mayor capacidad de discriminación. Estas apreciaciones coinciden con los hallazgos previos de Meissner y Kassin (2002). Recordemos que los autores sitúan al entrenamiento en la base del efecto del sesgo del investigador, al haber encontrado, tanto en un pequeño meta-análisis que incluyen al principio de su artículo como en el experimento que describen a continuación, que los programas de entrenamiento no producen mejoras en las tasas de precisión, sino que aumentan las probabilidades de juzgar una declaración como falsa.

Bull (2004) hace una revisión de los resultados de 12 estudios distintos sobre el entrenamiento en detección del engaño. Aunque en la mayoría de los casos los participantes eran estudiantes, en algunos habían participado muestras de profesionales. A partir de los datos observados, el autor ofrece la siguiente conclusión:

Los estudios revisados parecen, al menos superficialmente, contradecirse unos a otros respecto al entrenamiento en indicios de engaño. Mientras algunos han encontrado que éste mejoró la ejecución, otros no (Bull, 2004, p. 259).

Respecto a aquellos estudios en los que se había tratado de entrenar a policías, el autor concluye:

Deben superarse obstáculos cruciales para que el entrenamiento tenga una oportunidad de ser efectivo. Éstos tienen que ver con pedir a quienes van a ser entrenados (es decir, los policías) en la detección del engaño a partir de indicios no verbales:

1. Que dejen de basarse en creencias falsas sobre los indicios.
2. Que estén dispuestos a reconocer que los investigadores en este tema disponen de información importante y relevante.
3. Que estén dispuestos a aceptar que algunos de ellos (especialmente los que tienen más experiencia) no son buenos en esta tarea y que por lo tanto habrán cometido errores cruciales en el pasado (Bull, 2004, pp. 265-266).

En definitiva, podemos concluir que: (a) aunque hay algunas discrepancias entre los distintos estudios, en general el entrenamiento parece producir una leve mejora en la precisión global de los no profesionales; (b) sin embargo esto no refleja una mayor capacidad de discriminación ni de verdades ni de mentiras, sino un sesgo de respuesta consistente en clasificar más declaraciones como falsas (Meissner y Kassin, 2002); (c) en algunos estudios con policías el resultado ha ido en la dirección opuesta a lo esperado; es decir, el entrenamiento parece reducir las tasas de precisión de los policías, y sin embargo parece aumentar la confianza en sus juicios.

UNA POSIBLE EXPLICACIÓN DEL EFECTO DEL ENTRENAMIENTO: LA INFLUENCIA DE DISTINTAS MODALIDADES DE ENTRENAMIENTO SOBRE LOS JUICIOS DE VERACIDAD

Una razón que podría explicar porqué el entrenamiento sesga los juicios hacia la mentira podría ser la propia naturaleza de los programas de entrenamiento que se ofrecen. Tanto el entrenamiento que reciben los profesionales durante su formación, como el que se ofrece en los manuales, como el que se ha utilizado en los experimentos de laboratorio tienen como objetivo la detección de *mentiras*, y se dirige el foco atencional de los entrenandos exclusivamente hacia la búsqueda de indicios de engaño. Como resultado, al finalizar el entrenamiento y confrontar al individuo entrenado con una tarea de evaluación de la veracidad, éste va a buscar activamente la presencia de los indicios en los que ha sido entrenado (indicios de engaño, no de verdad), y va a ser especialmente sensible a su presencia. No obstante, ya se ha señalado anteriormente que los indicios de engaño son de naturaleza probabilística; es decir, aunque pueden aparecer más a menudo cuando una persona miente, eso no significa que no puedan estar presentes en algunos casos en los que se está diciendo la verdad.

La sensibilidad del individuo entrenado a la aparición de los indicios sobre los que ha recibido entrenamiento y la búsqueda activa de los mismos, podría llevarle no solamente a encontrarlos cuando la otra persona miente, sino también cuando dice la verdad. Al percibir dichos indicios, el detector podría rápidamente inferir que el emisor está mintiendo, sin preocuparse de buscar otros elementos de contraste que pudieran sugerirle que dice la verdad. Esta estrategia resultaría en muchos juicios de mentira, lo que a su vez daría lugar a una elevada precisión al juzgar declaraciones falsas, pero a una baja precisión al juzgar las declaraciones verdaderas. Esto explicaría, además, porqué la precisión global aumenta levemente sólo en algunos casos, mientras que en otros permanece constante.

Si todo este razonamiento es correcto, entonces un programa de entrenamiento centrado sobre la detección de la verdad (en vez del engaño) y en el que se enfatizaran los indicadores conductuales de verdad (en vez de los de la mentira) debería resultar en un aumento de juicios de verdad y en un mayor número de aciertos al juzgar verdades que al juzgar mentiras.

Estos planteamientos se han sometido a comprobación empírica en los dos estudios que presentamos a continuación.

ESTUDIO 5

Diseño e hipótesis

En este estudio se establecieron dos grupos que recibirían dos formas diferentes de entrenamiento, así como un tercer grupo (control) que no sería entrenado:

a. *Grupo de entrenamiento Tradicional (GR-TRAD)*. Se enseñó a los observadores a identificar claves supuestamente indicadoras de mentira.

b. *Grupo de entrenamiento Alternativo (GR-ALT)*. Se enseñó a los observadores a identificar claves supuestamente indicadoras de verdad. Éstas eran las conductas opuestas a las claves empleadas en el entrenamiento tradicional.

c. *Grupo control (GR-CONT)*. Los participantes asignados a este grupo no recibieron ningún tipo de entrenamiento. En su lugar, se les dio una charla sobre el Efecto Pigmalión.

Se empleó un diseño mixto (Tabla 33). En un primer momento (pre-entrenamiento), todos los participantes juzgaron la veracidad de un conjunto de declaraciones grabadas en vídeo, indicando asimismo la confianza en sus juicios. Esto permitió establecer su línea base. Seguidamente, un conjunto de ellos recibió el entrenamiento tradicional, otro el entrenamiento alternativo y un tercer grupo, de control, se sometió a una sesión en la que se impartieron unos contenidos ajenos al tema del engaño. Posteriormente (post-entrenamiento), todos los participantes tuvieron que hacer una nueva tarea de juicio de

veracidad similar a la del pre-entrenamiento en la que, además de expresar sus juicios y la confianza en los mismos, tuvieron que indicar el grado en que habían detectado los indicadores en los que habían sido entrenados.

Tabla 33. *Esquema General del Diseño Experimental.*

Grupos		Pre-entrenamiento	Entrenamiento	Post-entrenamiento
Entrenamiento Tradicional (GR-TRAD)	Subgrupo 1	Vídeo A	Para detectar mentiras Claves de la mentira	Vídeo B
	Subgrupo 2	Vídeo B		Vídeo A
Entrenamiento Alternativo (GR-ALT)	Subgrupo 1	Vídeo A	Para detectar verdades Claves de la verdad	Vídeo B
	Subgrupo 2	Vídeo B		Vídeo A
Control (GR-CONT)	Subgrupo 1	Vídeo A	Sin Entrenamiento (contenidos ajenos al tema)	Vídeo B
	Subgrupo 2	Vídeo B		Vídeo A

Las hipótesis específicas del estudio fueron las siguientes:

Hipótesis 1: El porcentaje de juicios de mentira de quienes se sometían al entrenamiento tradicional (GR-TRAD) será mayor en el post-entrenamiento que en el pre-entrenamiento.

Hipótesis 2: El porcentaje de juicios de mentira de quienes se sometían al entrenamiento alternativo (GR-ALT) será menor en el post-entrenamiento que en el pre-entrenamiento.

No se esperaba ninguna variación entre el pre-entrenamiento y el post-entrenamiento en el porcentaje de juicios de mentira de los participantes del grupo control (GR-CONT).

Hipótesis 3: Los participantes de ambos grupos entrenados (GR-TRAD y GR-ALT) mostrarán mayor confianza en sus juicios después del entrenamiento que antes del mismo.

No se esperaba ninguna variación entre el pre-entrenamiento y el post-entrenamiento en la confianza de los participantes del grupo control.

Las hipótesis 4 y 5 se refieren a las percepciones de los participantes sobre los indicios en los que habían sido entrenados. Como explicaremos detalladamente más adelante, se utilizaron los mismos indicios para el GR-TRAD y el GR-ALT, pero en sus direcciones opuestas. Por ejemplo, si para el GR-TRAD un incremento en una determinada conducta se asoció con la

mentira, para el GR-ALT un decremento en esa misma conducta se asoció con la verdad. Nuestras hipótesis sobre las claves fueron las siguientes:

Hipótesis 4: Los participantes sobre-estimarán la frecuencia de aparición de los indicios en los que hayan sido entrenados.

Hipótesis 5: La frecuencia percibida de aparición de los indicios será un buen predictor de los juicios de los participantes.

Preparación del material

Para el desarrollo del experimento se precisó del siguiente material:

(a) Vídeos evaluación: Son los dos vídeos que se emplearon en las fases de pre-entrenamiento y post-entrenamiento.

(b) Cuestionario de claves. Se utilizaron para determinar las claves (indicadores conductuales de la mentira o de la verdad) presentes en los posibles vídeos evaluación y para decidir cuáles de ellas deberían incluirse en el entrenamiento.

(c) Programas de entrenamiento. Éstos incluyeron una exposición oral que se apoyó con diapositivas en *PowerPoint* y dos tipos de vídeos: vídeos ejemplo, en los que una actriz ejemplificó las claves en las que se estuviera entrenando a los participantes, y vídeos ejercicio, en los que aparecía una persona hablando con naturalidad. Los vídeos ejercicio se utilizaron durante el entrenamiento para que los participantes practicasen la identificación de las claves en las que se les estaba entrenando y determinaran si lo narrado era cierto o no.

(d) Cuestionarios pre-entrenamiento y post-entrenamiento, para juzgar la veracidad de las declaraciones.

Se expone a continuación cómo se elaboró todo este material.

Selección de las claves para el entrenamiento

Para la selección de las claves para el entrenamiento se acudió en primer lugar al meta-análisis de DePaulo, Lindsay, et al. (2003) sobre los indicadores

reales del engaño. Adicionalmente se consultaron los artículos sobre entrenamiento publicados, para comprobar qué claves se habían utilizado. Las claves que se seleccionaran debían reunir las siguientes características:

a) En primer lugar debían ser “no diagnósticas”, es decir, no tenían que haber mostrado en otros estudios una relación significativa con la verdad ni con la mentira. Este requisito se planteó porque lo que se pretendía mostrar en este estudio es que el entrenamiento puede sesgar la respuesta sin que varíe la precisión; el empleo de claves diagnósticas podría afectar dicha precisión.

b) Debían ser fácilmente identificables. Es decir, tenían que ser indicios que se pudieran identificar sin necesidad de ningún tipo de aparataje adicional.

c) Tenían que poder explicarse con claridad en sus dos formas inversas (por ejemplo, cantidad de parpadeos vs. tiempo sin parpadear). Esto es importante porque, por razones de control experimental, era necesario que, a excepción del tipo de entrenamiento, el resto de variables y condiciones fuesen las mismas en todos los grupos. Fue por esto que se decidió utilizar las mismas claves, pero en sus formas opuestas, en ambas modalidades de entrenamiento.

Teniendo en cuenta estos requisitos, se hizo una pre-selección de posibles claves verbales y no-verbales.

Elaboración de los vídeos evaluación

Para elaborar los vídeos evaluación, se partió inicialmente los 48 *clips* de vídeo ya grabados que se utilizaron en la investigación previa sobre el efecto de la experiencia (Estudio 3). Se seleccionaron 20 de estos *clips* y se editaron para que tuvieran una duración similar (entre 57 y 70 segundos). Entonces se montaron en dos vídeos, A y B, cuidando que: (a) cada uno de tales vídeos contuviera 10 declaraciones: cinco verdaderas y cinco falsas; (b) ninguno de los declarantes que apareciera en el Vídeo A volviera a aparecer en el Vídeo B; (c) la cantidad de hombres y mujeres que mentían y decían la verdad fuera la

misma en ambos vídeos; y (d) el orden de las declaraciones en cada vídeo fuera aleatorio.

En la Tabla 34 se presentan las características de los vídeos A y B. La duración media de las declaraciones fue de 65 segundos, tanto para el Vídeo A como para el Vídeo B. Se visualizaron ambos vídeos para estimar informalmente con qué frecuencia las claves pre-seleccionadas se mostraban en sus dos direcciones, y se redujo así el número de claves relevantes a una cifra manejable (ocho).

Tabla 34. *Composición y Características de los Vídeos A y B.*

Orden de presentación	Vídeo A			Vídeo B		
	Valor de verdad	Sexo del emisor	Duración (segundos)	Valor de verdad	Sexo del emisor	Duración (segundos)
1	Verdad	Mujer	70	Mentira	Hombre	58
2	Verdad	Mujer	67	Mentira	Mujer	67
3	Verdad	Hombre	59	Verdad	Mujer	69
4	Mentira	Mujer	70	Mentira	Hombre	69
5	Mentira	Hombre	67	Mentira	Mujer	69
6	Verdad	Hombre	63	Verdad	Mujer	56
7	Mentira	Mujer	64	Mentira	Mujer	68
8	Verdad	Mujer	70	Verdad	Hombre	70
9	Mentira	Hombre	62	Verdad	Hombre	63
10	Mentira	Mujer	58	Verdad	Mujer	60

Elaboración de los cuestionarios para la selección de claves

Se elaboraron dos cuestionarios de percepción de claves (Cuestionario de claves 1 y Cuestionario de claves 2). Cada uno de ellos recogía cuatro de las ocho claves inicialmente seleccionadas, y contenía escalas que iban de -2 a 2 mediante las cuales se evaluaba el grado en que cada testigo en los vídeos evaluación mostraba las conductas relevantes seleccionadas. Dos ejemplos de tales escalas se muestran en la Figura 12.

Sujeto 1

Seriedad			Sonrisas			Superficialidad			Detalles		
-2	-1	0	1	2	-2	-1	0	1	2		

Figura 12. *Ejemplo de dos de las Escalas de los Cuestionarios de Claves.*

Participantes

Un total de 89 estudiantes de psicología (78 mujeres y 11 hombres; edad: $M = 20.98$, $DE = 1.34$) cumplimentó estos cuestionarios. Hubo 41 (36 mujeres y 5

hombres; edad: $M = 20.90$, $DE = 1.22$) que primero vieron el Vídeo B mientras cumplimentaban el Cuestionario de claves 2, y después vieron el Vídeo A mientras cumplimentaban el Cuestionario de claves 1. Otros 48 observadores (42 mujeres y 6 hombres; edad: $M = 21.04$, $DE = 1.45$) vieron, en primer lugar, el Vídeo A y cumplimentaron el Cuestionario 2, y después vieron el Vídeo B y cumplimentaron el Cuestionario 1.

Procedimiento

A los participantes se les daban las siguientes instrucciones (el ejemplo variaba según se tratara del cuestionario 1 ó 2):

“A continuación vas a ver diez fragmentos de vídeo. En cada uno aparece una persona narrando unos hechos. Tu tarea consiste en cumplimentar, para cada una de las diez personas, las cuatro escalas referentes a su conducta que se muestran más abajo. Por ejemplo, si te parece que la persona está muy seria debes señalar el -2, si te parece que está algo seria el -1, si no te parece que esté ni seria ni sonriente el 0, si te parece que sonríe o se ríe algo el 1, y si te parece que sonríe o ríe mucho el 2. Después de visualizar a cada sujeto dispondrás de unos segundos para cumplimentar las cuatro escalas correspondientes a ese sujeto antes de pasar a visualizar el siguiente”.

Con el fin de seleccionar las claves definitivas, se introdujeron los datos en el programa SPSS y se analizaron separadamente para cada vídeo y cuestionario. Una puntuación media de una clave cercana a 0 indicaría que ninguna de sus dos formas opuestas es percibida notoriamente más que la otra. Si bien en la mayoría de los casos las puntuaciones medias diferían significativamente de 0, ya fuera en un vídeo o en el otro, lo cierto es que, en general y en términos absolutos, no se apartaban mucho de esta cifra. Se estableció, por lo tanto, un intervalo alrededor de 0, cuyos límites fueran ± 0.50 , y se decidió que las puntuaciones de las claves que se deberían seleccionar tenían que caer dentro de este intervalo en ambos vídeos, siendo además significativamente distintas tanto de -0.50 como de 0.50. Esto se comprobó con una serie de pruebas t que determinaron la conveniencia de seleccionar cuatro conductas: (a) discurso continuo vs. correcciones del lenguaje, (b) discurso ininterrumpido vs. pausas, (c) fluidez del habla vs. repeticiones de palabras y frases, y (d) movilidad de las cejas vs. inmovilidad de las cejas (Tabla 35).

Tabla 35. Datos de las Cuatro Claves Seleccionadas Para el Entrenamiento.

Claves			Vídeo A						Vídeo B			
	<i>Media</i>	<i>DE</i>	Contraste: -0.50		Contraste: 0.50		<i>Media</i>	<i>DE</i>	Contraste: -0.50		Contraste: 0.50	
			<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>			<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Discurso continuo Correcciones del lenguaje	-0.10	0.38	6.73	<.001	-10.19	<.001	0.14	0.37	11.99	<.001	-6.76	<.001
Discurso ininterrumpido Pausas	0.37	0.41	14.60	<.001	-2.21	.032	0.15	0.40	10.20	<.001	-5.49	<.001
Fluidez del habla Repeticiones palabras frases	-0.10	0.52	5.32	<.001	-7.98	<.001	-0.04	0.46	7.54	<.001	-6.36	<.001
Movilidad de las cejas Inmovilidad de las cejas	-0.04	0.44	7.27	<.001	-8.59	<.001	-0.37	0.40	2.16	.037	-13.87	<.001

Los extremos de cada una de las claves se definieron así:

(a) *Discurso continuo vs. correcciones del lenguaje*

Discurso continuo. El hablante narra su declaración de forma continua, sin detenerse para corregir o sustituir palabras. Este extremo se asoció a la verdad.

Correcciones del lenguaje. Correcciones de una o varias palabras dentro de una misma frase, ya sea:

- porque después de haber dicho una palabra completa el hablante se da cuenta de que no es la apropiada e inmediatamente la corrige,
- o porque el hablante ha empezado a decir una palabra y en mitad de la palabra se da cuenta de que no es la adecuada, dejándola incompleta y sustituyéndola por la adecuada.

Este extremo se asoció a la mentira

(b) *Discurso ininterrumpido vs. pausas*

Discurso ininterrumpido. Las palabras y frases aparecen de forma natural, continuada, sin cortes ni espacios de silencio innecesarios, con un ritmo y una cadencia naturales y adecuados para una buena comprensión. Este extremo se asoció a la verdad.

Pausas. Se caracteriza por la presencia de pausas de uno o varios de los tipos siguientes:

- Pausas vacías. Pausas durante las que el hablante guarda silencio.
- Pausas llenas. Pausas durante las cuales el hablante emite vocalizaciones como “ah”, “uhm”, “er”, “hmm”.
- Pausas mezcladas. Pausas compuestas de una parte vacía y otra llena.

Este extremo se asoció a la mentira.

(c) Fluidez del habla vs. repeticiones de palabras o frases.

Fluidez del habla. El discurso es fluido, sin repeticiones de palabras o frases innecesarias. Este extremo se asoció a la verdad.

Repeticiones de palabras o frases. El hablante repite palabras, frases, o fragmentos de palabras o de frases, pero no porque haya pausas, errores del discurso o correcciones. Este extremo se asoció a la mentira.

(d) Movilidad de las cejas vs. inmovilidad de las cejas

Movilidad de las cejas. Durante el discurso el sujeto realiza muchos movimientos ascendentes o descendentes de las cejas. Este extremo se asoció a la verdad.

Inmovilidad de las cejas. Durante el discurso las cejas del sujeto permanecen mucho tiempo en su lugar, sin moverse. Este extremo se asoció a la mentira.

Elaboración de los programas de entrenamiento

A partir de este punto se elaboraron los programas de entrenamiento. Como se ha indicado, hubo dos modalidades de entrenamiento: una centrada en las claves asociadas a la mentira y otra centrada en las claves asociadas a la verdad, con la particularidad de que las claves fueron las mismas en ambos entrenamientos, pero en sentido opuesto. Ambos programas debían ser idénticos, salvo el énfasis en la verdad y sus indicadores en uno y en la mentira y los suyos en el otro. La sesión de instrucción para el grupo control se diseñó para que tuviera una duración y estructura (definiciones, ejemplos, vídeos, etc.) similar a la de los entrenamientos.

Al diseñar los programas de entrenamiento se cumplió con los requisitos que, según Frank y Feeley (2003), debe tener un buen entrenamiento (particularmente aquellos relevantes para los propósitos del presente proyecto, a saber, examinar si diferentes tipos de entrenamiento tienen un efecto

diferencial sobre los juicios de verdad y de mentira). La instrucción, pues, consistiría en una sesión de no menos de una hora de duración (Frank y Feeley, 2003), que incluiría una charla ayudada por una presentación en *PowerPoint*. Además, junto a la definición, explicación y ejemplificación de cada una de las claves por parte del entrenador, se iban introduciendo oportunamente vídeos (los que denominamos “vídeos ejemplo”) en los que aparecía una persona mostrando la clave explicada, tanto en una sola frase, como a lo largo de una declaración completa.

Elaboración de los vídeos ejemplo

La ejemplificación de las claves durante el entrenamiento fue un aspecto sobre el que se trabajó especialmente, con el fin de que quedara claro para los observadores en qué consistía cada clave. Por lo tanto, una vez que se determinó cuales serían las claves, se estableció un proceso de selección de dos actores (un hombre y una mujer) para que las ejemplificaran⁵⁶.

Para la selección de los actores, se realizó una búsqueda a través de Internet de compañías teatrales locales y provinciales que contasen con actores y actrices semiprofesionales. Posteriormente, se contactó telefónicamente o vía correo electrónico con cada compañía, y se propuso a sus miembros la posibilidad de participar en un proceso de selección para la grabación de los vídeos ejemplo. Se organizó una prueba de selección de candidatos. Se citó a los actores y se les explicó en qué consistía el trabajo y las condiciones de contratación.

Entonces se dio paso a la prueba de selección propiamente dicha. Se entregó un texto a cada actor o actriz y se le dejó unos minutos para que se familiarizase con él. Posteriormente se le pidió que leyese el texto tratando de interpretarlo, indicándole, en cada lectura sucesiva, que fuese introduciendo distintos ejemplos de cada una de las claves que posteriormente, en caso de ser seleccionado, tendría que representar. Para ello se le describía

⁵⁶ Se decidió contar con actores porque estos pueden repetir la misma representación de forma idéntica, variando únicamente la conducta de interés. Ekman sostiene además que los actores disponen de un gran control conductual, con lo que son capaces de realizar a voluntad determinados gestos y movimientos y de inhibir otros.

detalladamente cada clave. Finalmente, un actor y una actriz fueron seleccionados. Se contactó con los candidatos restantes para agradecerles su participación y comunicarles que no habían sido seleccionados.

Se instruyó minuciosamente a los actores sobre qué decir y hacer con el fin de ejemplificar las claves seleccionadas, y se grabó su actuación representando cada clave individual una sola vez en una sola frase (fácilmente identificable) e inserta en una declaración en la que apareciese repetidamente en lugares pertinentes (ver Anexo N). Los actores aparecían siempre sentados en una silla, de cuerpo entero. El investigador se situó detrás de la cámara, de forma que no era visible en la imagen. Se editaron tales grabaciones, creando así Vídeos ejemplo que se mostraron a los participantes que se sometieron al entrenamiento. Puesto que el entrenamiento acabó siendo más extenso de lo inicialmente previsto, se decidió finalmente prescindir de los Vídeos ejemplo de uno de los actores. Se utilizaron, pues, sólo los vídeos de la actriz, por resultar sus actuaciones ligeramente más satisfactorias que las del actor.

Elaboración de los vídeos ejercicio

El objetivo de los vídeos ejercicio era que los participantes tuviesen la oportunidad de practicar, durante el proceso de entrenamiento, la identificación de cada una de las claves. Era necesario, por lo tanto, disponer de un conjunto de *clips* en los que, tal como ocurría en los Vídeos pre-entrenamiento y post-entrenamiento, apareciesen una serie de personas haciendo declaraciones.

Se reclutaron alumnos de primer curso del Título Propio de Ciencias de la Seguridad de la Universidad de Salamanca, a quienes se ofreció una bonificación en la nota final de la asignatura Psicología de la Delincuencia si participaban. Se convocó a quienes accedieron y se les pidió que cumplimentaran un cuestionario sobre sus actitudes acerca de una serie de temas polémicos de actualidad en aquel momento (ver Anexo O). A continuación, tuvieron que hacer una declaración en la que se reflejara su verdadera actitud y otra contraria a la misma. El tema de dicha declaración fue uno de aquellos incluidos en el cuestionario sobre los cuales el participante tuviera una actitud más extrema. Las declaraciones de los participantes se

grabaron en vídeo. Cada uno de ellos aparecía sentado, de cuerpo entero. El entrevistador quedaba fuera del encuadre.

De las declaraciones obtenidas de este modo se seleccionaron dos, una verdadera (reflejaba la actitud real del hablante) y la otra falsa (contra-actitudinal), para ser utilizadas como Vídeos ejercicio. El emisor de cada una de estas declaraciones fue distinto, tratándose en una ocasión de un hombre y en la otra de una mujer. No se usaron más declaraciones con el fin de evitar la fatiga de los participantes, ya que el entrenamiento iba a ser largo.

Elaboración de los cuestionarios pre-entrenamiento y post-entrenamiento

Se diseñaron los cuestionarios que los observadores deberían cumplimentar al juzgar las declaraciones en las fases de pre-entrenamiento y post-entrenamiento. En ellos tenían que indicar lo siguiente para cada narrador:

a) Si éste decía la verdad o mentía. En los cuestionarios del GR-TRAD, la pregunta clave era: “¿El sujeto miente?”. Los participantes tenían que señalar la palabra “Sí” o “No” que aparecía después de dicha pregunta. En los cuestionarios del GR-ALT la pregunta clave era: “¿El sujeto dice la verdad?”. Por último, los participantes del GR-CONT tenían que señalar simplemente las palabras “Verdad” o “Mentira”. En todos los casos, el orden de las opciones de respuesta se contrabalanceó para cada muestra.

b) La confianza o seguridad en cada juicio de veracidad en una escala tipo *Likert* que iba de 1 (*totalmente inseguro*) a 7 (*totalmente seguro*).

c) En los cuestionarios del post-entrenamiento, los participantes también tenían que señalar el grado en que habían percibido, en la conducta del emisor, las claves en las que habían sido entrenados. Esto tenía por objeto hacer que realmente utilizaran tales claves, así como comprobar si sobreestimaban la presencia vs. ausencia de las mismas según la modalidad de entrenamiento. Esta comprobación era necesaria para constatar que los cambios en los juicios obedecen realmente al efecto del entrenamiento. En las Figuras 13 y 14 se presentan ejemplos de las preguntas del cuestionario que medían la

percepción de las claves por parte de los participantes en los grupos de entrenamiento.

SUJETO 1																			
¿En qué medida has percibido las siguientes claves?																			
Pausas					Inmovilidad de las cejas					Correcciones del lenguaje				Repeticiones de palabras y Frases					
0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
En consecuencia, ¿el sujeto miente? SI NO																			
Totalmente inseguro							Totalmente seguro												
1	2	3	4	5	6	7													

Figura 13. *Pregunta Para Medir la Percepción de las Claves en el Post-Entrenamiento Usada con el GR-TRAD.*

SUJETO 1																			
¿En qué medida has percibido las siguientes claves?																			
Discurso ininterrumpido					Movilidad de las cejas					Discurso continuo				Fluidez del habla					
0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
En consecuencia, ¿el sujeto dice la verdad? SI NO																			
Totalmente inseguro							Totalmente seguro												
1	2	3	4	5	6	7													

Figura 14. *Pregunta Para Medir la Percepción de las Claves en el Post-Entrenamiento Usada con el GR-ALT*

Método

Participantes

Participaron en este estudio 115 estudiantes universitarios (90 mujeres y 25 hombres; edad: $M = 20.99$, $DE = 1.82$), todos ellos alumnos de tercero de Psicología en la Universidad de Salamanca, que estaban cursando la asignatura de Psicología Social. Se presentaron voluntariamente para la

realización de esta investigación, recibiendo a cambio una bonificación en su nota final en la mencionada asignatura. Treinta y seis participantes conformaron el GR-TRAD, 40 el GR-ALT y 39 el GR-CONT. Las pequeñas desigualdades entre los grupos se deben a que algunos participantes no se presentaron el día del experimento. La asignación de participantes a cada una de las modalidades de entrenamiento se realizó siguiendo la premisa de mantener, en la medida de lo posible, la media de edad y la (des)proporción de mujeres y hombres constante en los tres grupos.

Procedimiento

A. Fase de pre-entrenamiento

Cada uno de los tres grupos de participantes fue citado el mismo día (para evitar en la medida de lo posible cualquier efecto de filtración de información entre los grupos), aunque en franjas horarias diferentes. Los participantes debían acudir a la Facultad de Psicología de la Universidad de Salamanca. Cada subgrupo dentro de cada grupo (véase la Tabla 33) debía acudir a un aula distinta. Se habían habilitado, por lo tanto, dos aulas, con un investigador en cada una de ellas. Los participantes de un aula (Aula A) verían el Vídeo A en esta fase, mientras que los participantes del otro aula (Aula B) verían el Vídeo B.

A medida que los participantes iban entrando en el aula, el investigador les iba entregando una tarjeta con un número. Debían conservar dicha tarjeta durante todo el experimento, ya que indicaba el número de asiento que debían ocupar en las aulas y figuraba en los dos cuestionarios que cumplimentarían. Esto permitiría emparejar los dos cuestionarios de cada persona.

Los participantes encontraron sobre los pupitres un impreso que debían cumplimentar expresando el carácter voluntario de su participación en el estudio, las instrucciones de la tarea y los cuestionarios de respuesta. El investigador leyó en primer lugar el impreso de participación voluntaria (Anexo P) y todos los participantes lo cumplimentaron y firmaron. A continuación se les explicó que su tarea consistiría en ver 10 declaraciones grabadas en vídeo y juzgar la veracidad de cada una de ellas. A los del GR-TRAD se les dijo que

debían indicar si cada declaración era mentira o no, a los del GR-ALT que debían indicar si cada declaración era verdad o no, y a los del GR-CONT si cada declaración era mentira o verdad. Cada juicio se haría inmediatamente al finalizar la declaración correspondiente, momento en el que se detendría la proyección durante unos segundos. Los juicios debían señalarse en el cuestionario. Además de anotar su juicio de veracidad, los participantes tenían que indicar su confianza en dicho juicio.

El investigador leyó las instrucciones de cumplimentación del cuestionario (Anexo Q), y tras preguntar a los participantes si tenían alguna duda sobre la tarea comenzó con la proyección de las declaraciones. Estas se reprodujeron con un ordenador portátil al que se habían conectado unos altavoces externos y un cañón que proyectaba la imagen sobre una pantalla situada en la parte frontal del aula. Al finalizar la tarea, el investigador recogió los cuestionarios e indicó a los participantes que a continuación se daría inicio a la siguiente fase del estudio. La duración de esta primera fase fue, aproximadamente, de 30 minutos.

B. Fase de Entrenamiento

El entrenamiento tuvo, en sí mismo, tres fases diferentes. En un primer momento, se hizo una introducción general. Se trataba, por un lado, de motivar a los participantes indicándoles la importancia de saber con precisión cuando alguien miente (o dice la verdad, según se tratara del GR-TRAD o del GR-ALT). Además, se hacía referencia a la dificultad intrínseca de evaluar con precisión la veracidad. También se aludía a que las personas confían mucho en claves erróneas y que, no obstante, existen otras claves más útiles en las que se les iba a entrenar.

Una vez introducido el programa de entrenamiento, se entraba en una segunda fase, en la que a cada grupo se le instruía en las claves correspondientes siguiendo un protocolo establecido: (a) el entrenador definía la primera clave de entrenamiento (la definición aparecía, además, en una transparencia), la explicaba y la ejemplificaba él mismo; (b) el entrenador dictaba la definición a los participantes que estaban siendo entrenados; (c) el

entrenador mostraba un breve *clip* de vídeo en el cual una actriz decía una única frase mostrando la clave (vídeo ejemplo de una frase), seguido de otro *clip* en el que decía la misma frase sin dicha clave (vídeo ejemplo de una frase) para que se apreciara la diferencia (Anexo N); (d) el entrenador mostraba vídeos más largos para ejemplificar la clave (Vídeos ejemplo de una declaración):

- El GR-TRAD veía un vídeo en el que la actriz narraba unos hechos (ver Anexo N) mostrando la clave, se señalaban las apariciones de la misma y se decía que si alguien hablaba de ese modo era más probable que mintiera que si hablaba como la persona del siguiente vídeo que se iba a presentar. Se mostraba entonces un vídeo en el cual la actriz narraba los mismos hechos, con las mismas palabras, pero sin mostrar ninguna de las cuatro claves del GR-TRAD.
- Al GR-ALT se le decía que nadie habla mostrando la clave constantemente (por ej., con un discurso absolutamente continuo, o ininterrumpido, o fluido, o moviendo continuamente las cejas) aunque diga la verdad. Veían entonces un vídeo en el que la actriz narraba unos hechos mostrando la clave todo el tiempo (hablaba de manera continua [sin ninguna corrección], ininterrumpida [sin pausas], fluida [sin repeticiones] o moviendo todo el tiempo las cejas) y el entrenador enfatizaba la artificialidad del mismo. Se resaltaba entonces que lo que tenían que hacer era advertir si había fragmentos durante los cuales el hablante mostrara la clave (por ej., fragmentos en los que hablara de forma continua). Se les mostraba entonces el mismo vídeo que se había mostrado al GR-TRAD para ejemplificar la forma opuesta de la clave, pero se focalizaba su atención hacia aquellas partes del discurso en las cuales la actriz hablaba de manera continua, ininterrumpida, fluida o moviendo las cejas.

Finalmente, el entrenador invitaba a los entrenandos a que, una vez más, releyeran la definición de la clave. El proceso se repetía con el resto de claves.

En este punto se hacía un receso de 10 minutos para que los participantes pudieran descansar. Tras el descanso se entraba en la tercera parte del entrenamiento, que consistía en presentar dos vídeos a modo de ejercicio (vídeos ejercicio). En cada uno de ellos aparecía una persona haciendo una declaración, y los participantes debían inferir su veracidad en función de las claves en las que habían sido entrenados. Después de cada vídeo, el entrenador preguntaba a los participantes quiénes consideraban que la declaración era verdad y quiénes consideraban que era mentira. Entonces:

- Al GR-TRAD le indicaba que la declaración era mentira, y centraba su atención en aquellas formas de las claves que aparecían en la declaración en las que habían sido entrenados, con el fin de que, en base a la aparición de las mismas, realmente asumieran que la declaración era falsa.
- Al GR-ALT, le indicaba que la declaración era verdad, y centraba su atención en aquellas formas de las claves que aparecían en la declaración en las que habían sido entrenados, con el fin de que, en base a la aparición de las mismas, realmente asumieran que la declaración era verdadera.

En realidad, ambos grupos de entrenamiento vieron los dos mismos vídeos, uno verdadero y otro falso.

El objetivo de este ejercicio era hacer que los participantes se centrasen exclusivamente en las formas de las claves en las que habían sido entrenados, ignorando cualquier otro indicio. Con ello se pretendía que, en la tarea de evaluación de la credibilidad de la fase post-entrenamiento, atendieran de forma selectiva a estas claves e interpretasen la declaración en consecuencia, es decir, de forma sesgada hacia la verdad o la mentira, ignorando otros posibles indicios que pudieran aparecer en la declaración.

La duración total aproximada del entrenamiento fue de 1 hora y 40 minutos (incluyendo el descanso).

El GR-CONT no recibió ningún tipo de entrenamiento. En su lugar, se le dio una charla sobre el efecto Pigmalión. Se adoptó un formato de presentación similar (en la medida de lo posible) al empleado con los grupos experimentales. La duración de esta charla también fue parecida a la de los programas de entrenamiento.

C. Fase de post-entrenamiento

El investigador dividió de nuevo a los participantes de cada grupo en los dos subgrupos originales del pre-entrenamiento y los ubicó en aulas distintas. El subgrupo que durante el pre-entrenamiento había juzgado el Vídeo A en el Aula A fue enviado ahora al Aula B para juzgar el Vídeo B, y viceversa. La tarea y las instrucciones fueron idénticas a las de la fase de pre-entrenamiento, con la salvedad de que se indicó explícitamente a los participantes de los grupos de entrenamiento que debían basar sus juicios en las claves en las que se les había entrenado. Para ello se emplearon unos cuestionarios ligeramente distintos a los del pre-entrenamiento, ya que se incluyeron las claves del entrenamiento y una escala junto con cada una de ellas en la que debían indicar, antes de hacer el juicio de veracidad, en qué medida habían percibido la presencia de la misma. La escala abarcaba de 0 (*clave ausente*) a 4 (*clave muy presente*) (véanse las Figuras 13 y 14). Una vez concluida la tarea el investigador recogió los cuestionarios, agradeció su colaboración a los participantes y les indicó que durante una clase de la asignatura se les daría información sobre el estudio.

Resultados

Un Análisis de la Varianza (ANOVA) de un factor mostró que las medias de edad de los Grupos de Entrenamiento Tradicional, $M = 20.69$, $DE = 1.39$, Entrenamiento Alternativo, $M = 21.05$, $DE = 1.55$, y Control, $M = 21.21$, $DE = 2.35$, no fueron significativamente distintas, $F_{(2,112)} = 0.77$, $p = .467$. Respecto al sexo, en los tres grupos se mantuvo aproximadamente la misma proporción entre mujeres y hombres, $\chi^2(2, N = 115) = 2.83$, $p = .243$. La precisión global pre-entrenamiento (es decir, antes de la manipulación experimental) fue $M = 55.04$, $DE = 15.01$, y resultó superior al nivel de azar del 50%, $t_{(114)} = 3.60$, $p <$

.001. Con el fin de constatar que los estímulos empleados no eran atípicos, se comparó esta media con el valor 53.46%, correspondiente a la precisión media ponderada hallada por Bond y DePaulo (2006) en su meta-análisis sobre la precisión de los juicios de veracidad al que ya nos hemos referido a lo largo de esta Tesis. La diferencia no fue significativa, $t_{(114)} = 1.13$, $p = .260$). Esto sugiere que los resultados hallados no pueden deberse a idiosincrasias en los estímulos usados.

El porcentaje medio de juicios de mentira antes del entrenamiento fue $M = 49.91$, $DE = 11.74$. Un ANOVA de un factor mostró que el efecto del grupo (GR-TRAD vs. GR-ALT vs. GR-CONT) sobre el porcentaje de juicios de mentira antes del entrenamiento no era significativo, $F_{(2,112)} = 1.05$, $p = .352$, $\eta^2 = .018$. Esto significa que el enunciado de la pregunta sobre el juicio de credibilidad en los cuestionarios (que, como ya se ha indicado, era distinto en las distintas condiciones) no tuvo ninguna influencia sobre los juicios.

Análisis del sesgo en los juicios debido al entrenamiento

Se hizo un ANOVA mixto en el que se introdujeron el valor de verdad (veracidad real de las declaraciones) y momento (pre- vs. post-entrenamiento) como variables independientes de medidas repetidas, y el grupo de entrenamiento (GR-TRAD vs. GR-ALT vs. GR-CONT) como variable independiente intergrupo. La variable dependiente fue el porcentaje de juicios de mentira (en adelante PJM). Este análisis arrojó un efecto significativo de la variable valor de verdad, $F_{(1,112)} = 14.44$, $p < .001$, $\eta^2 = .114$, mostrando que los participantes efectuaron más juicios de mentira cuando las declaraciones eran realmente falsas ($M = 53.65$, $DE = 16.88$) que cuando eran verdaderas ($M = 46.61$, $DE = 16.00$). Esto refleja cierto grado de precisión en sus juicios.

El efecto principal de la variable grupo de entrenamiento también fue significativo, $F_{(2,112)} = 27.22$, $p < .001$, $\eta^2 = .327$, mostrando que el GR-TRAD efectuó el mayor porcentaje de juicios de mentira ($M = 60.83$, $DE = 12.22$), seguido del GR-CONT ($M = 46.67$, $DE = 11.26$) y del GR-ALT ($M = 43.88$, $DE = 8.20$). Las pruebas *post-hoc* de *Scheffe* mostraron que las diferencias entre el GR-TRAD y el GR-ALT ($p < .001$) y entre el GR-TRAD y el GR-CONT ($p <$

.001) eran significativas, pero que la diferencia entre el GR-ALT y el GR-CONT no era significativa ($p = .508$) (ver Figura 15). El efecto principal de la variable momento no fue significativo, $F_{(1,112)} = 0.29$, $p = .589$, $\eta^2 < .001$.

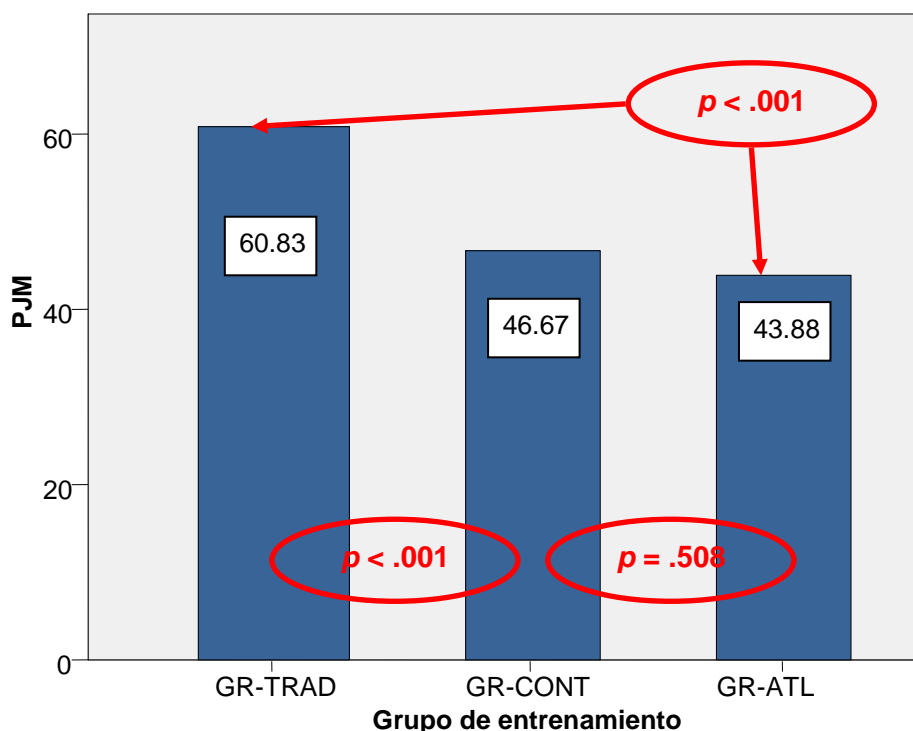


Figura 15. Porcentajes de Juicios de Mentira.

El efecto de la interacción Valor de Verdad x Grupo de Entrenamiento fue significativo, $F_{(2,112)} = 3.61$, $p = .030$, $\eta^2 = .061$. Aunque en todos los grupos se efectuaron más juicios de mentira cuando las declaraciones eran falsas que cuando eran verdaderas, los análisis *post-hoc* mostraron que esta diferencia no alcanzó significatividad para el GR-ALT (Tabla 36). Esto puede reflejar la influencia del entrenamiento. En el GR-ALT, y a diferencia de los otros grupos, éste se centró sobre claves de la verdad, lo cual pudiera haber reducido el porcentaje de juicios de mentira, especialmente al juzgar mentiras. Pero este efecto debería ser más pronunciado en el post-entrenamiento. Efectivamente, mientras que en el pre-entrenamiento el GR-ALT obtuvo un PJM marginalmente mayor al juzgar mentiras ($M = 52.50$, $DE = 16.76$) que al juzgar verdades ($M = 44.50$, $DE = 20.5$; $F_{(1,112)} = 2.82$, $p = .096$, $\eta^2 = .025$), en el post-entrenamiento dicho grupo obtuvo un PJM marginalmente menor al juzgar

mentiras ($M = 36.00$, $DE = 18.78$) que al juzgar verdades ($M = 42.50$, $DE = 18.78$; $F_{(1,112)} = 2.94$, $p = .089$, $\eta^2 = .026$). Sin embargo, la interacción Valor de Verdad x Grupo de Entrenamiento x Momento no alcanzó el nivel de significación estadística, $F_{(2,112)} = 1.67$, $p = .193$, $\eta^2 = .029$.

Tabla 36. *Porcentaje Medio (Desviaciones Típicas) de Juicios de Mentira en Función de la Modalidad de Entrenamiento y la Veracidad de las Declaraciones.*

	Verdades	Mentiras
Entrenamiento Tradicional (GR-TRAD)*	56.94 (13.90)	64.72 (16.99)
Entrenamiento Alternativo (GR-ALT)	43.50 (13.69)	44.25 (12.59)
Control (GR-CONT)**	40.26 (15.64)	53.08 (14.72)

* Diferencias significativas con $p < .050$

** Diferencias significativas con $p < .001$

El efecto de la interacción Grupo de Entrenamiento x Momento fue significativo, $F_{(2,112)} = 21.48$, $p < .001$, $\eta^2 = .277$. En concreto, en línea con la Hipótesis 1, en el GR-TRAD se produjo un incremento en el PJM como consecuencia del entrenamiento (véase Figura 16). Un contraste *post-hoc* realizado mediante el subcomando EMMEANS de SPSS indicó que este incremento es estadísticamente significativo, $F_{(1,112)} = 30.37$, $p < .001$, $\eta^2 = .213$. Además, en línea con la Hipótesis 2, en el GR-ALT el porcentaje de juicios de mentira disminuyó (Figura 16). Un contraste *post-hoc* indicó que esta diferencia es estadísticamente significativa, $F_{(1,112)} = 9.73$, $p = .002$, $\eta^2 = .080$. Por último, en el GR-CONT el porcentaje de juicios de mentira fue, aparentemente, menor en la tarea post-entrenamiento que en la tarea pre-entrenamiento (Figura 16). Sin embargo, un contraste *post-hoc* mostró que la diferencia entre ambas no alcanza la significación estadística, $F_{(1,112)} = 2.92$, $p = .090$, $\eta^2 = .025$.

Como ya hemos indicado anteriormente, el efecto de la interacción Grupo de Entrenamiento x Momento x Valor de Verdad no fue significativo, lo cual indica que el efecto de la interacción Grupo x Momento se mantuvo constante para las declaraciones verdaderas y falsas.

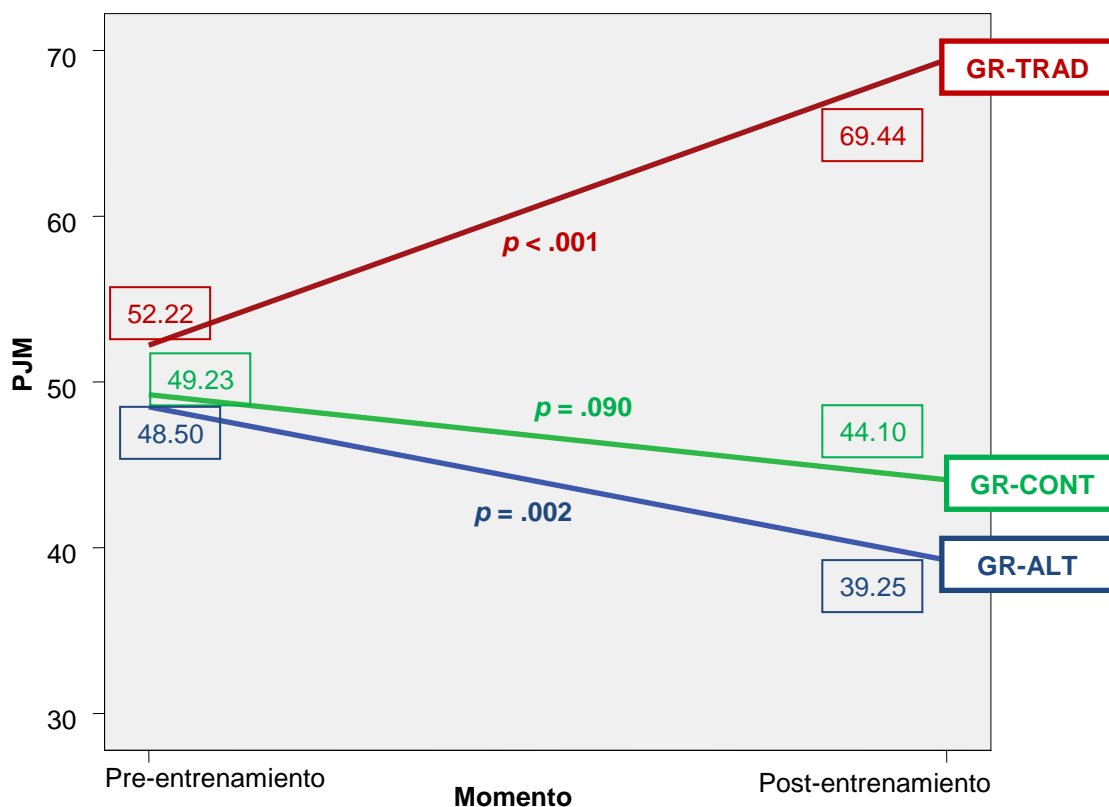


Figura 16. Porcentaje de Juicios de Mentira Emitidos por Cada Grupo en Cada Momento.

Estos últimos resultados prestan apoyo a las dos primeras hipótesis. Sin embargo, para mostrar que los juicios de los observadores realmente reflejan un sesgo de respuesta, se hizo un análisis adicional sobre B'' . En el GR-TRAD las señales eran las mentiras, porque el entrenamiento se centró sobre la mentira, mientras que siguiendo la misma lógica, en el GR-ALT las señales eran las verdades. No obstante, con el fin de poder comparar los resultados de los distintos grupos se tomó como referencia la mentira, codificando los aciertos y las falsas alarmas de los tres grupos tal como correspondería al GR-TRAD, y calculando a continuación B'' . Así, se consideró que una mentira identificada como tal era un acierto, mientras que una verdad juzgada como mentira era una falsa alarma, con independencia del grupo. Para calcular B'' se empleó la fórmula de Donaldson (1992). Los valores de B'' pueden abarcar desde -1 a 1. Una respuesta sin sesgo arrojaría un valor de $B'' = 0$, los valores negativos reflejan un sesgo a juzgar las declaraciones como falsas y los valores positivos reflejan un sesgo a juzgar las declaraciones como verdaderas. Se esperaba que, en comparación con el pre-entrenamiento, en el post-

entrenamiento el valor de B'' se hubiera reducido en el GR-TRAD, hubiera aumentado en el GR-ALT y no hubiera experimentado cambios significativos en el GR-CONT. Asimismo, en el post-entrenamiento se esperaba un valor de B'' significativamente menor que 0 en el GR-TRAD, significativamente mayor que 0 en el GR-ALT y cercano a 0 en el GR-CONT.

Se hizo pues un ANOVA mixto para examinar la influencia del momento (pre- vs. post-entrenamiento) y del grupo de entrenamiento (GR-TRAD vs. GR-ALT vs. GR-CONT) sobre los valores de B'' . El efecto del momento no fue significativo, $F_{(1,112)} = 0.19$, $p = .663$, $\eta^2 = .002$, pero sí lo fue el efecto del grupo de entrenamiento, $F_{(2,112)} = 26.86$, $p < .001$, $\eta^2 = .324$. La prueba de *Scheffe* mostró que el valor medio de B'' en el GR-TRAD ($M = -.33$, $DE = .36$) era significativamente menor que en el GR-ALT ($M = .22$, $DE = .30$; $p < .001$) y que en el GR-CONT ($M = .12$, $DE = .36$; $p < .001$). La diferencia entre el GR-ALT y el GR-CONT no fue significativa ($p = .498$). Este efecto principal del grupo se vio no obstante matizado por una interacción Momento x Grupo de Entrenamiento, $F_{(2,112)} = 18.74$, $p < .001$, $\eta^2 = .251$. En consonancia con la Hipótesis 1, en el GR-TRAD se produjo un descenso en el valor medio de B'' como consecuencia del entrenamiento (pre-entrenamiento, $M = -.07$, $DE = .48$; post-entrenamiento, $M = -.59$, $DE = .43$). Un contraste *post-hoc* indicó que este descenso era estadísticamente significativo, $F_{(1,112)} = 26.04$, $p < .001$, $\eta^2 = .189$. Además, en consonancia con la Hipótesis 2, en el GR-ALT el valor medio de B'' se incrementó (pre-entrenamiento, $M = .07$, $DE = .42$; post-entrenamiento, $M = .36$, $DE = .47$). Un contraste *post-hoc* indicó que esta diferencia era estadísticamente significativa, $F_{(1,112)} = 8.91$, $p = .003$, $\eta^2 = .074$. Por último, en el GR-CONT el valor medio de B'' se incrementó (pre-entrenamiento, $M = .05$, $DE = .55$; post-entrenamiento, $M = .20$, $DE = .51$). Sin embargo, un contraste *post-hoc* mostró que dicho incremento no alcanzaba la significación estadística, $F_{(1,112)} = 2.56$, $p = .113$, $\eta^2 = .022$.

Se hicieron análisis adicionales que mostraron que en el pre-entrenamiento ningún grupo estaba sesgado, ya que en ningún caso fueron significativas las diferencias entre sus valores promedios de B'' y 0 (GR-TRAD: $t_{(35)} = -0.86$, $p = .396$; GR-ALT: $t_{(39)} = 1.07$, $p = .291$; GR-CONTR: $t_{(38)} = 0.65$, $p = .521$). Sin embargo, en el post-entrenamiento el GR-TRAD estaba significativamente

sesgado hacia juicios de mentira, $t_{(35)} = -8.20$, $p < .001$, y el GR-ALT hacia juicios de verdad, $t_{(39)} = 4.88$, $p < .001$. Por su parte, el GR-CONT estaba ligera aunque significativamente sesgado hacia juicios de verdad, $t_{(38)} = 2.47$, $p = .018$.⁵⁷

Impacto del entrenamiento sobre la confianza en los juicios

Se hizo un ANOVA mixto para examinar el efecto del valor de verdad de las declaraciones, momento (pre- vs. post-Entrenamiento) y el grupo de entrenamiento (GR-TRAD vs. GR-ALT vs. GR-CONT) sobre la confianza en los juicios. Se halló un efecto principal significativo del valor de verdad, $F_{(1,112)} = 7.43$, $p = .007$, $\eta^2 = .062$, que indicaba que, tomando globalmente todas las puntuaciones, se observaba una mayor confianza para las verdades, $M = 4.32$, $DE = 0.89$, que para las mentiras, $M = 4.17$, $DE = 0.95$.

El efecto principal de la variable Momento también fue significativo, $F_{(1,112)} = 19.11$, $p < .001$, $\eta^2 = .146$. La confianza post-entrenamiento ($M = 4.36$, $DE = 0.98$) fue superior a la confianza pre-entrenamiento ($M = 4.12$, $DE = 0.90$). Sin embargo, este efecto se vio matizado por una interacción Momento x Grupo de Entrenamiento, $F_{(2,112)} = 6.87$, $p = .002$, $\eta^2 = .109$. Como se observa en la Figura 17, la confianza aumentó como resultado del entrenamiento tanto en el GR-TRAD (confianza media pre- y post-entrenamiento: $M = 4.23$, $DE = 0.78$; $M = 4.72$, $DE = 0.91$; este incremento fue significativo según las pruebas *post-hoc*, $F_{(2,112)} = 22.76$, $p < .001$, $\eta^2 = .169$) como en el GR-ALT ($M = 4.10$, $DE = 0.98$; $M = 4.40$, $DE = 0.90$, $F_{(2,112)} = 9.16$, $p = .003$, $\eta^2 = .076$), pero no varió significativamente en el GR-CONT ($M = 4.03$, $DE = 0.92$; $M = 4.00$, $DE = 1.01$

⁵⁷ También se calculó un ANOVA análogo sobre A' , que es un estimador insesgado de la precisión derivado de la Teoría de la Detección de Señales (Green y Swets, 1966; Stanislaw y Todorov, 1999; Swets et al., 2000). Se utilizó para ello la fórmula de Rae (1976). A' va de 0 a 1. El valor 0.5 corresponde a una precisión equivalente al azar y el valor 1 a una precisión perfecta. El único efecto significativo fue un débil efecto principal del grupo, $F_{(2,112)} = 3.14$, $p = .047$, $\eta^2 = .053$. La prueba *post-hoc* de Scheffe mostró que el GR-ALT ($M = .51$, $DE = .15$) y el GR-TRAD ($M = .56$, $DE = .15$) no diferían significativamente ($p = .301$), como tampoco lo hacían ($p = .692$) el GR-TRAD y el GR-CONT ($M = .59$, $DE = .14$), pero la precisión del GR-CONT fue marginalmente mayor que la del GR-ALT ($p = .051$). En cualquier caso, ni el momento ni la interacción Momento x Grupo fueron significativos. En otras palabras, aunque el entrenamiento sesgó las respuestas, no incrementó ni redujo la precisión para ningún grupo. En el pre-entrenamiento el valor promedio de A' fue $M = .57$, $DE = .22$; en el post-entrenamiento fue $M = .54$, $DE = .19$. Aunque en términos absolutos el valor fuera muy bajo, en ambos casos estuvo significativamente por encima del nivel de azar, respectivamente, $t_{(114)} = 3.47$, $p = .001$; $t_{(114)} = 2.00$, $p = .048$.

$F_{(2,112)} = 0.10$, $p = .759$, $\eta^2 = .00$). No hubo ningún otro efecto significativo en este ANOVA.

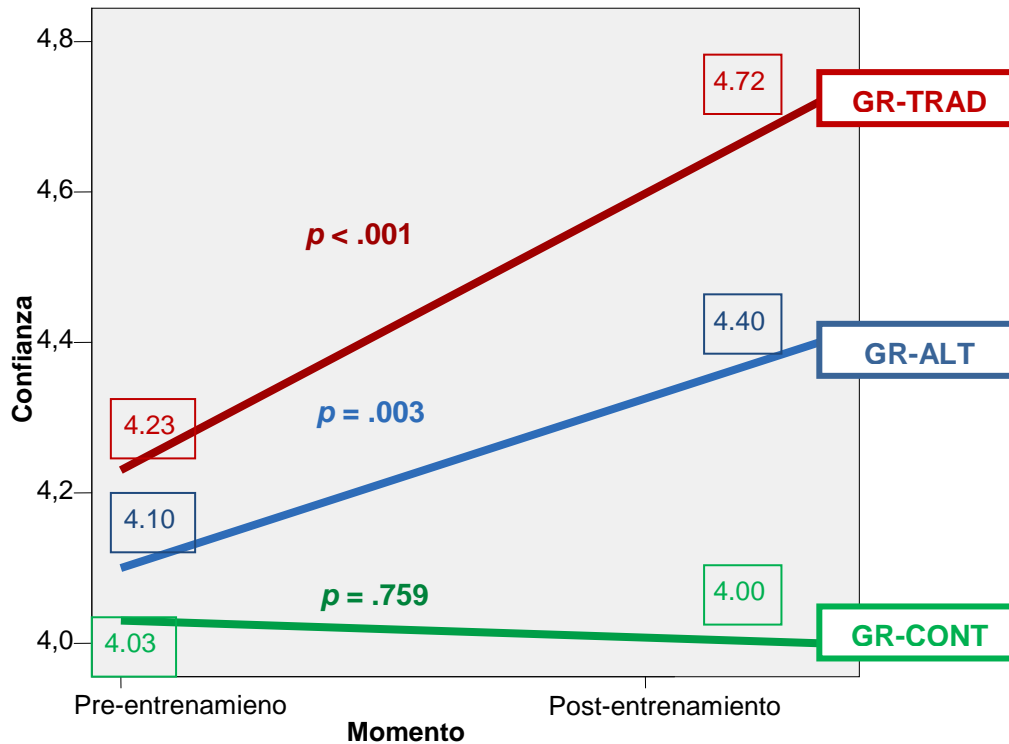


Figura 17. *Confianza de Cada Grupo en los Juicios en Cada Momento.*

Percepción de las claves del entrenamiento

La Hipótesis 4 predecía que los participantes sobreestimarían la frecuencia con la que las claves se mostrarían en la dirección en la que hubieran sido entrenados. Las estimaciones numéricas de los participantes sobre la presencia de las claves en los *clips* del vídeo del post-entrenamiento se codificaron de manera que mayores valores equivalían a más correcciones del lenguaje (discurso menos continuo), más pausas (discurso menos ininterrumpido), más repeticiones (discurso menos fluido) y mayor inmovilidad de las cejas (menor movilidad de cejas). A continuación, se calculó el valor promedio para cada una de estas cuatro variables a través de los 10 juicios hechos por cada participante en el post-entrenamiento. Los valores podían estar entre 0 y 4. Si la modalidad de entrenamiento influye sobre la percepción de las claves del modo hipotetizado, entonces esperaríamos que las medias en

estas variables fueran significativamente mayores en el GR-TRAD que en el GR-ALT.

Se calculó, pues, un Análisis Múltiple de la Varianza (MANOVA) para contrastar esta hipótesis. La variable independiente fue el grupo de entrenamiento (GR-TRAD vs. GR-ALT) y las variables dependientes los valores promedios de las cuatro claves. A nivel multivariado, el análisis resultó significativo, *Lambda de Wilks* = .341, $F_{(4,71)} = 34.30$, $p < .001$, $\eta^2 = .569$. Los análisis univariados fueron significativos para tres de las claves empleadas. En comparación con los participantes del GR-ALT, los del GR-TRAD percibieron más correcciones (GR-ALT, $M = 1.48$, $DE = 0.53$; GR-TRAD, $M = 1.76$, $DE = 0.59$; $F_{(1,74)} = 4.97$, $p = .029$, $\eta^2 = .063$), más pausas (GR-ALT, $M = 1.37$, $DE = 0.46$; GR-TRAD, $M = 2.55$, $DE = 0.56$; $F_{(1,74)} = 101.98$, $p < .001$, $\eta^2 = .580$) y más repeticiones (GR-ALT, $M = 1.49$, $DE = 0.48$; GR-TRAD, $M = 1.84$, $DE = 0.68$; $F_{(1,74)} = 6.92$, $p = .010$, $\eta^2 = .085$). Las estimaciones sobre el movimiento de cejas del GR-ALT ($M = 2.18$, $DE = 0.56$) no difirieron significativamente de las del GR-TRAD ($M = 1.99$, $DE = 0.54$; $F_{(1,74)} = 2.47$, $p = .120$, $\eta^2 = .032$). En definitiva, la modalidad del entrenamiento tuvo un efecto significativo sobre la percepción de tres de las cuatro claves.

Una última cuestión de interés es si la percepción de las claves sería un buen predictor de los juicios de credibilidad, tal como planteaba la Hipótesis 5. Es decir, ¿la tendencia a clasificar las declaraciones como falsas fue mayor cuantas más correcciones, pausas, repeticiones e inmovilidad de las cejas se percibían? Para examinar esta cuestión se hicieron un total de 10 análisis de regresión logística mediante el método *introducir*. Cada análisis se refería a uno de los diez juicios hechos por los participantes en la fase de post-entrenamiento (cinco en los que se juzgaron declaraciones verdaderas y otros cinco en los que se juzgaron declaraciones falsas). Las cuatro claves conductuales (correcciones, pausas, repeticiones e inmovilidad de las cejas) se introdujeron como variables predictoras, y el juicio (verdad vs. mentira) como variable dependiente.

Según se observa en la Tabla 37, las estimaciones de los observadores sobre la presencia de las claves permitían pronosticar con notable acierto sus

juicios. Así, el porcentaje medio total de juicios correctamente pronosticados por las ecuaciones logísticas fue del 87.76% (rango 81.58%-93.42%), siendo del 86.11% para las verdades (rango 73.33%-94.59%) y del 87.43% para las mentiras (rango 77.78%-93.44%). También se puede observar que el χ^2 del modelo, que examina la diferencia entre el valor de $-2LL$ para el modelo sólo con la constante y con las cuatro variables predictoras, es significativo en todos los casos. Además, los valores del R^2 de Nagelkerke fueron muy altos, lo que indica que los modelos logísticos explicaron un porcentaje sustancial de varianza. En conclusión, los datos muestran que, sin duda, los juicios de los participantes estuvieron guiados por su percepción de las claves.

Tabla 37. *Porcentaje de Juicios Correctamente Pronosticados por las Ecuaciones Logísticas Basadas en las Estimaciones de las Claves, χ^2 del Modelo y R^2 de Nagelkerke.*

Clip	Juicios de Verdad	Juicios de Mentira	Global	χ^2 (4, N = 76)	R^2 de Nagelkerke
Verdad 1	91.84	77.78	86.84	56.07*	.717
Verdad 2	92.11	92.11	92.11	70.27*	.804
Verdad 3	80.95	82.35	81.58	44.02*	.588
Verdad 4	81.82	86.05	84.21	48.15*	.629
Verdad 5	73.33	93.44	89.47	49.31*	.758
Mentira 1	91.89	92.31	92.11	83.19*	.887
Mentira 2	84.62	86.49	85.53	49.53*	.639
Mentira 3	80.77	92.00	88.16	58.96*	.746
Mentira 4	89.19	79.49	84.21	45.70*	.603
Mentira 5	94.59	92.31	93.42	79.81*	.867
<i>Media</i>	86.11	87.43	87.76		

Nota: Verdad 1, Verdad 2, etc.: primera, segunda, etc. declaración verdadera juzgada por los observadores, con independencia del vídeo.

* $p < .001$

Conclusiones

En este estudio se examinó el efecto de dos modalidades distintas de entrenamiento sobre los juicios de veracidad emitidos en una tarea posterior de evaluación de la credibilidad. Las cinco hipótesis que se plantearon recibieron apoyo empírico. Así, en comparación con la fase previa al entrenamiento, los participantes hicieron más juicios de mentira después de haber sido entrenados para detectar el engaño (Hipótesis 1) y más juicios de verdad después de haber sido entrenados para detectar la verdad (Hipótesis 2). Parece ser, por lo tanto, que tras el entrenamiento los participantes se implicaban en un proceso sesgado de búsqueda de información, buscando activamente las claves en las que habían sido entrenados y desechando las demás. Es importante señalar

que ambos grupos entrenados fueron expuestos a los mismos vídeos en la fase final. Sin embargo, indicaron haber percibido claves opuestas según hubieran sido asignados al grupo tradicional o al grupo alternativo (Hipótesis 4). Además, la percepción de los indicios tuvo un impacto sobre los juicios (Hipótesis 5), que se hicieron con mucha confianza (Hipótesis 3). Por lo tanto, los programas de entrenamiento sesgan las decisiones de los observadores e incrementan su confianza.

Es importante señalar que en el estudio que se acaba de describir, el cuestionario que rellenaron los participantes de los grupos entrenados en la fase post-entrenamiento difería del cuestionario de la fase pre-entrenamiento y del que cumplimentaron los participantes del grupo control. En ambos grupos experimentales se pedía, para cada juicio y previa emisión de dicho juicio, una estimación numérica referida a la percepción de la presencia de cada una de las claves de entrenamiento. Esto no podía hacerse en la fase pre-entrenamiento, ni podía hacerse con el grupo control puesto que éste no había recibido ningún entrenamiento. Específicamente, los cuestionarios post-entrenamiento de los grupos entrenados contenían una serie de cuatro escalas que los participantes tenían que cumplimentar para cada emisor antes de expresar el juicio de credibilidad para dicho emisor. Las escalas medían el grado en que los participantes habían percibido las claves en las que habían sido entrenados. Nuestro objetivo a la hora de incluir dichas escalas fue (a) recoger datos para comprobar las Hipótesis 4 y 5, y (b) asegurarnos de que los participantes hacían sus juicios basándose en las claves en las que habían sido entrenados y no en otra información o impresiones generales. Sin embargo, no podemos saber si los cambios en los juicios de mentira de los grupos entrenados son el resultado de los programas de entrenamiento, de haber tenido que completar las escalas, o una combinación de ambos elementos. Ciertamente, al pedir a los participantes de los grupos entrenados que evaluaran la presencia de las claves, les obligamos, de forma sistemática, a hacer juicios basados en dichas claves. Esta estrategia cognitiva pudo haber aumentado los efectos del programa de entrenamiento (es decir, si los participantes entrenados no hubieran sido obligados a procesar cuidadosamente las claves del entrenamiento, entonces podrían haber

mostrado cambios menos drásticos en sus juicios), o pudo incluso haber sido suficiente para causar las diferencias observadas. El Estudio 6 se realizó para comprobar las Hipótesis 1 y 2 sin obligar a los participantes de los grupos de entrenamiento a escudriñar cuidadosamente la conducta de los emisores en busca de las claves de entrenamiento. Si aún así los datos confirman las hipótesis, entonces es que los programas de entrenamiento pueden sesgar las respuestas de los entrenandos por sí mismos, independientemente de si se fuerza a los participantes a escudriñar la conducta del emisor o no.

ESTUDIO 6

Objetivo, hipótesis y diseño

El objetivo del Estudio 6 fue contrastar las Hipótesis 1, 2 y 3 del Estudio 5, pero utilizando para la tarea pre y post-entrenamiento el mismo cuestionario para el GR-ALT, el GR-TRAD y el GR-CONT. En concreto, se suprimió en los cuestionarios de los dos grupos experimentales la parte en la que los participantes debían expresar en una escala en qué medida habían percibido las claves. De este modo, no se forzaba a los participantes a emitir sus juicios de veracidad en función de la percepción previa de las claves en las que habían sido entrenados. Al mantener el resto de condiciones constante entre los dos grupos experimentales, cualquier diferencia observada en los resultados se debería al tipo de entrenamiento recibido. El diseño y fases del Estudio 6 fue exactamente el mismo que en el Estudio 5.

Método

Participantes y Procedimiento

Los participantes fueron 124 estudiantes (112 mujeres y 12 hombres) que cursaban tercero de Psicología en la Universidad de Salamanca. La media de edad fue $M = 21.09$, $DE = 2.52$. Ninguno de los participantes de este estudio había tomado parte en el Estudio 5. A cambio de su participación, todos ellos recibieron una bonificación en sus notas en la asignatura de Psicología Social. El GR-TRAD contó con 46 participantes, el GR-ALT con 43 y el GR-CONT con 38. Estas pequeñas diferencias en los tamaños muestrales de los grupos se

produjeron porque algunas personas convocadas para participar en el experimento no se presentaron. El procedimiento de este estudio fue idéntico al empleado en el Estudio 5.

Resultados

Un ANOVA de un factor mostró que no había diferencias significativas en las medias de edad de los tres grupos (para el GR-TRAD, $M = 21.00$ años, $DE = 1.73$; para el GR-ALT, $M = 20.86$ años, $DE = 1.66$; y para el GR-CONT, $M = 21.45$ años, $DE = 3.79$), $F_{(2,121)} = 0.59$, $p = .559$. Además, la proporción de hombres y mujeres fue similar en los tres grupos. La precisión global pre-entrenamiento fue $M = 56.21\%$, $DE = 14.12$, significativamente superior al 50% esperado por azar $t_{(123)} = 4.90$, $p < .001$, y similar a la precisión global del 55.04% obtenida en el Estudio 5, $t_{(123)} = 0.92$, $p = .358$.

El PJM antes del entrenamiento fue $M = 49.27$, $DE = 12.95$. Un ANOVA mostró que el grupo (GR-TRAD vs. GR-ALT vs. GR-CONT) no tuvo una influencia significativa sobre el PJM previo al entrenamiento, $F_{(2,121)} = 0.90$, $p = .411$, $\eta^2 = .015$, lo cual indica que el enunciado de la pregunta sobre el juicio de veracidad en los distintos cuestionarios no afectó a los juicios de los participantes.

Análisis del sesgo en los juicios debido al entrenamiento

Al igual que en el Estudio 5, se realizó un ANOVA mixto en el que se introdujeron el valor de verdad de las declaraciones y el momento (pre- vs. post-entrenamiento) como variables independientes de medidas repetidas, y el grupo de entrenamiento (GR-TRAD vs. GR-ALT vs. GR-CONT) como variable independiente intergrupo. La variable dependiente fue el porcentaje de juicios de mentira (PJM). Este análisis arrojó un efecto significativo de la variable valor de verdad que indicó que se hacían más juicios de mentira cuando las declaraciones eran realmente falsas ($M = 55.10$, $DE = 18.32$) que cuando eran verdaderas ($M = 48.17$, $DE = 15.72$), $F_{(1,121)} = 14.70$, $p < .001$, $\eta^2 = .108$. Este resultado refleja que la precisión global estuvo por encima del nivel del azar.

El efecto principal del momento también fue significativo, $F_{(1,121)} = 7.74$, $p = .006$, $\eta^2 = .060$. Se hicieron más juicios de mentira tras el entrenamiento ($M = 54.19$ $DE = 21.30$) que antes de él ($M = 49.27$ $DE = 12.95$). Además, el efecto del grupo de entrenamiento también fue significativo, $F_{(2,121)} = 18.39$, $p < .001$, $\eta^2 = .233$. En concreto, los participantes del GR-TRAD fueron los que emitieron el mayor porcentaje de juicios de mentira ($M = 60.58$, $DE = 14.65$), seguidos de los del GR-CONT ($M = 49.21$, $DE = 11.18$) y finalmente de los del GR-ALT ($M = 45.12$, $DE = 10.26$) (Figura 18). Unas pruebas *post-hoc* de Scheffe revelaron que las diferencias entre el GR-TRAD y el GR-CONT, y entre el GR-TRAD y el GR-ALT fueron significativas (ambas $ps < .001$), pero el PJM del GR-ALT no se diferenció significativamente del PJM del GR-CONT ($p = .326$).

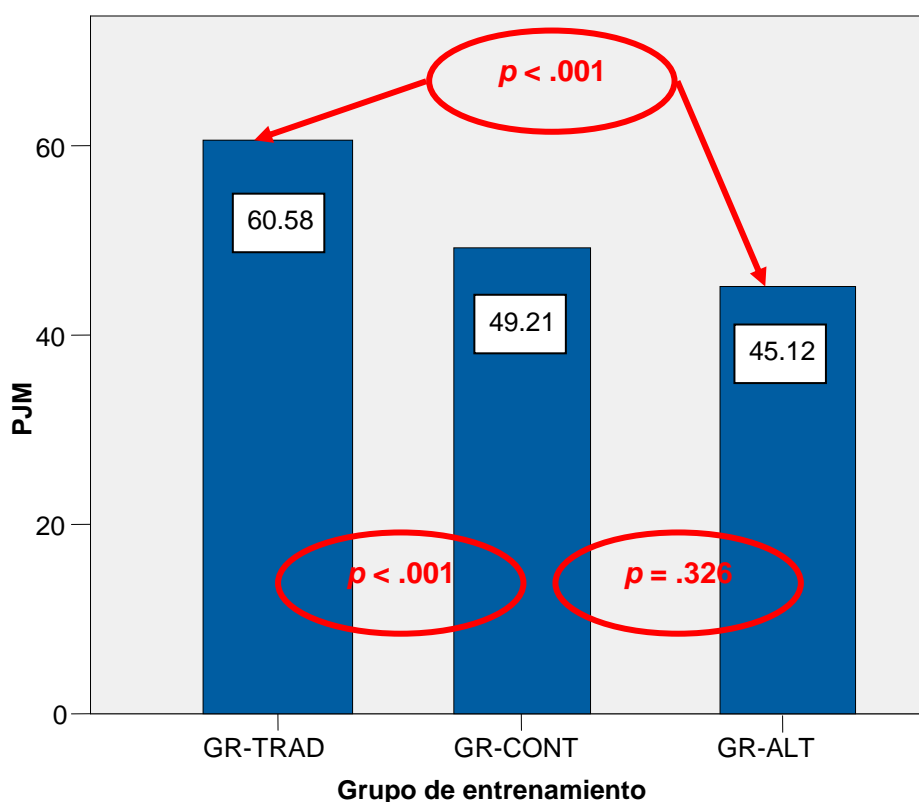


Figura 18. Porcentajes de Juicios de Mentira.

Se observó también una interacción significativa entre el Valor de Verdad y el Momento, $F_{(1,121)} = 8.14$, $p = .005$, $\eta^2 = .063$, la cual matizó el efecto principal del valor de verdad antes descrito. Las pruebas *post-hoc* realizadas para examinar los efectos simples mostraron que las diferencias en PJM entre declaraciones verdaderas y falsas fueron significativas antes del entrenamiento

($M = 55.35$, $DE = 19.89$, y $M = 43.09$, $DE = 18.40$, respectivamente), $F_{(1,121)} = 23.17$, $p < .001$, $\eta^2 = .161$, pero no después de él ($M = 54.85$, $DE = 26.46$, y $M = 53.26$, $DE = 25.21$, respectivamente), $F_{(1,121)} = 0.36$, $p = .550$, $\eta^2 = .003$. Esto sugiere que el entrenamiento redujo la precisión.

El efecto de la interacción Grupo de Entrenamiento x Momento también fue significativo, $F_{(2,121)} = 18.24$, $p < .001$, $\eta^2 = .232$, lo que matiza los efectos principales de estas variables antes expuestos. Las pruebas *post-hoc* mostraron que los participantes del GR-TRAD emitieron significativamente más juicios de mentira después del entrenamiento ($M = 69.77$, $DE = 20.64$) que antes del mismo ($M = 51.40$, $DE = 14.57$), $F_{(1,121)} = 38.94$, $p < .001$, $\eta^2 = .243$, mientras que los participantes del GR-ALT emitieron significativamente menos juicios de mentira después del entrenamiento ($M = 41.86$, $DE = 17.36$) que antes de él ($M = 48.37$, $DE = 9.74$), $F_{(1,121)} = 4.89$, $p = .029$, $\eta^2 = .039$. Estos resultados ofrecen apoyo a las Hipótesis 1 y 2. Finalmente, en el GR-CONT no se observaron diferencias significativas entre el PJM después ($M = 50.53$, $DE = 14.51$) y antes del entrenamiento ($M = 47.89$, $DE = 14.17$), $F_{(1,121)} = 0.71$, $p = .402$, $\eta^2 = .006$ (Figura 19). La interacción entre las tres variables independientes no fue significativa, lo que indica que la interacción Grupo x Momento se mantenía tanto para declaraciones falsas como verdaderas.

Al igual que en el Estudio 5, para observar si los juicios de los observadores realmente reflejan un sesgo de respuesta, se hizo un ANOVA adicional en el que se introdujeron las variables momento y grupo de entrenamiento como factores y los valores de B'' como variable dependiente. Tanto el efecto principal del momento, $F_{(1,121)} = 6.39$, $p = .013$, $\eta^2 = .050$, como del grupo de entrenamiento, $F_{(2,121)} = 16.50$, $p < .001$, $\eta^2 = .214$, fueron significativos. En general, los valores de B'' disminuyeron como resultado del entrenamiento (antes del entrenamiento: $M = .02$, $DE = .47$; después del entrenamiento: $M = -.13$, $DE = .65$). Unas pruebas *post-hoc* de *Scheffe* realizadas a continuación mostraron que el valor medio de B'' fue significativamente menor para el GR-TRAD ($M = -.32$, $DE = .39$) que para el GR-ALT ($M = .17$, $DE = .39$), $p < .001$, y que para el GR-CONT ($M = .00$, $DE = .40$), $p = .002$. Sin embargo, la diferencia entre el GR-ALT y el GR-CONT no fue significativa ($p = .175$).

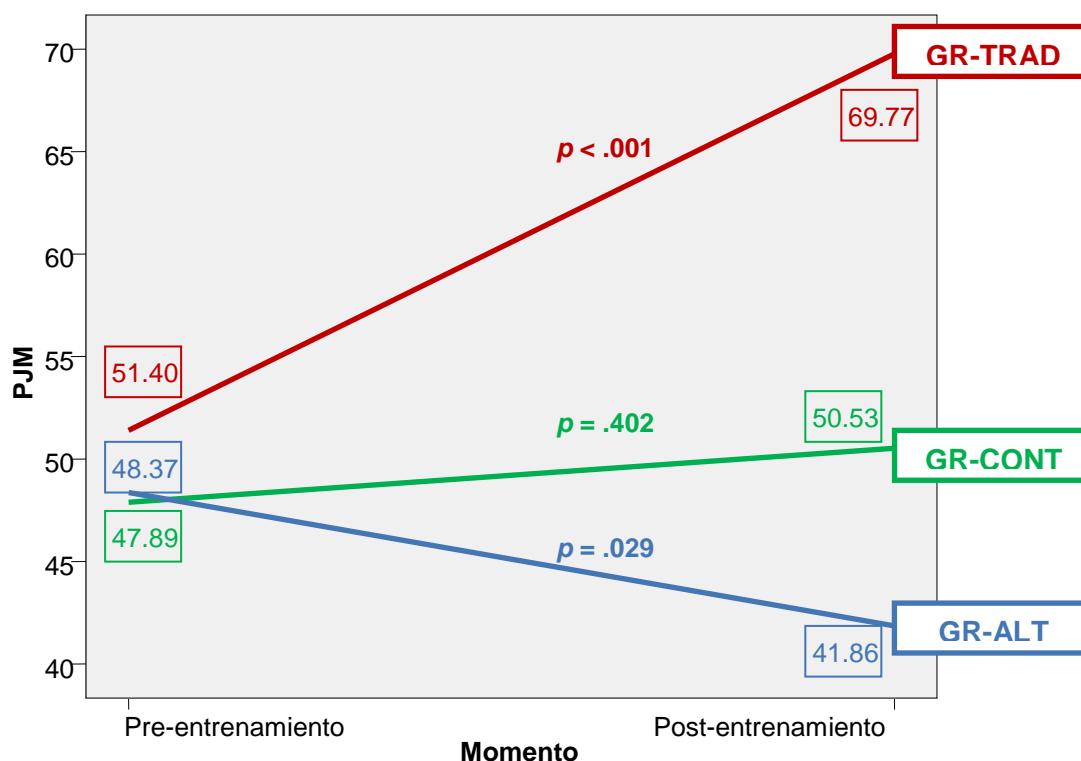


Figura 19. Porcentaje de Juicios de Mentira Emitidos por Cada Grupo en Cada Momento.

Estos efectos principales se vieron no obstante matizados por una interacción significativa entre Momento y Grupo de Entrenamiento, $F_{(2,121)} = 12.14$, $p < .001$, $\eta^2 = .167$. En línea con la Hipótesis 1, el entrenamiento causó una disminución del valor medio de B'' para el GR-TRAD (antes del entrenamiento: $M = -.06$, $DE = .49$; después del entrenamiento: $M = -.58$, $DE = .53$). Una prueba *post-hoc* mostró que dicha disminución fue significativa, $F_{(1,121)} = 27.25$, $p < .001$, $\eta^2 = .184$. Además, y en línea con la Hipótesis 2, el entrenamiento causó un incremento del valor medio de B'' en el GR-ALT (antes del entrenamiento: $M = .08$, $DE = .41$; después del entrenamiento: $M = .25$, $DE = .60$). Una prueba *post-hoc* mostró que dicha disminución fue marginalmente significativa, $F_{(1,121)} = 2.85$, $p = .094$, $\eta^2 = .023$. Finalmente, no se observó ninguna variación significativa en el valor medio de B'' en el GR-CONT como efecto del entrenamiento (antes del entrenamiento: $M = .05$, $DE = .52$; después del entrenamiento: $M = -.05$, $DE = .52$), $F_{(1,121)} = 0.78$, $p = .378$, $\eta^2 = .006$.

Algunos análisis adicionales mostraron que ningún grupo emitió juicios sesgados antes del entrenamiento, ya que en ningún caso los valores medios de B'' se diferenciaron significativamente de 0, ($t_{(42)} = -0.73$, $p = .472$, para el

GR-TRAD; $t_{(42)} = 1.31$, $p = .199$, para el GR-ALT; y $t_{(37)} = 0.57$, $p = .574$ para el GR-CONT). Sin embargo, después del entrenamiento se observó un sesgo significativo hacia la mentira en el GR-TRAD, $t_{(42)} = -7.21$, $p < .001$, y un sesgo significativo hacia la verdad en el GR-ALT, $t_{(42)} = 2.75$, $p = .009$. No se encontraron diferencias significativas en los valores medios de B'' antes y después del entrenamiento para el GR-CONT, $t_{(37)} = -0.55$, $p = .585$.⁵⁸

Impacto del entrenamiento sobre la confianza en los juicios

Se realizó un ANOVA mixto para examinar el efecto del valor de verdad (declaraciones verdaderas vs. falsas), momento (pre- vs. post-entrenamiento) y grupo de entrenamiento (GR-TRAD vs. GR-ALT vs. GR-CONT) sobre la confianza en los juicios. Se observó un efecto principal significativo de la variable momento, indicando que en general los participantes mostraron más confianza en sus juicios después del entrenamiento ($M = 4.46$, $DE = 1.03$) que antes de él ($M = 4.26$, $DE = 0.85$), $F_{(1,121)} = 8.77$, $p = .004$, $\eta^2 = .068$. No obstante, una interacción significativa entre las variables Momento y Grupo de Entrenamiento, $F_{(2,121)} = 4.42$, $p = .014$, $\eta^2 = .068$, indicó que estas diferencias fueron significativas solamente en el GR-TRAD y en el GR-ALT. Efectivamente, las pruebas *post-hoc* mostraron que mientras que la confianza aumentó tras el entrenamiento en el GR-TRAD (antes del entrenamiento: $M = 4.10$, $DE = 0.90$; después del entrenamiento: $M = 4.47$, $DE = 1.25$), $F_{(1,121)} = 12.19$, $p = .001$, $\eta^2 = .092$, y en el GR-ALT (antes del entrenamiento: $M = 4.40$, $DE = 0.76$; después del entrenamiento: $M = 4.66$, $DE = 0.70$), $F_{(1,121)} = 6.04$, $p = .015$, $\eta^2 = .048$, esto no fue así en el GR-CONT, en el que los valores de la confianza no variaron significativamente (antes del entrenamiento: $M = 4.29$, $DE = 0.87$; después del entrenamiento $M = 4.22$, $DE = 1.03$), $F_{(1,121)} = 0.44$, $p = .507$, $\eta^2 = .004$ (Figura 20). Estos resultados ofrecen apoyo a la Hipótesis 3.

⁵⁸ Se calcularon también los valores medios de A' . Antes del entrenamiento, el valor medio de A' fue $M = .59$, $DE = .20$, significativamente mayor que el nivel del azar, $t_{(123)} = 4.70$, $p < .001$. Después del entrenamiento, el valor medio de A' fue $M = .51$, $DE = .22$, el cual no se diferenció significativamente del valor del azar, $t_{(123)} = 0.47$, $p = .639$. Se calculó un ANOVA para examinar la influencia de las variables momento y grupo de entrenamiento sobre los valores medios de A' . Sólo el efecto principal de la variable momento fue significativo, $F_{(1,121)} = 7.52$, $p = .007$, $\eta^2 = .059$, indicando que la precisión fue mayor antes del entrenamiento que después.

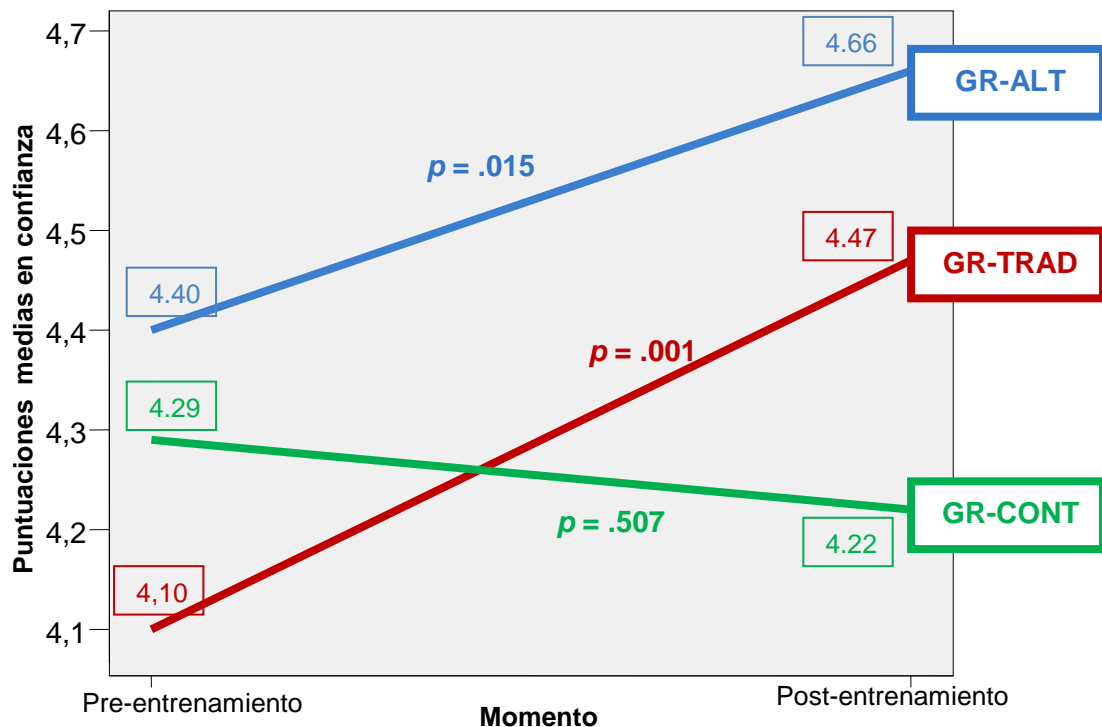


Figura 20. Porcentajes Medios de Confianza de Cada Grupo Antes y Después del Entrenamiento.

La interacción Valor de verdad x Grupo de Entrenamiento también fue significativa, $F_{(2,121)} = 3.37$, $p = .038$, $\eta^2 = .053$. Las pruebas *post-hoc* indicaron que los participantes del GR-ALT mostraron más confianza en sus juicios al juzgar declaraciones verdaderas ($M = 4.60$, $DE = 0.71$) que al juzgar declaraciones falsas ($M = 4.46$, $DE = 0.71$), aunque esta diferencia fue sólo marginalmente significativa, $F_{(1,121)} = 3.32$, $p = .071$, $\eta^2 = .027$. Sin embargo, esta tendencia no se observó ni en el GR-TRAD ni en el GR-CONT, $F_{s(2,121)} \leq 1.83$, $ps \geq .179$.

Conclusiones

El Estudio 6 replicó en general los resultados encontrados en el Estudio 5. Al no haber introducido en el cuestionario post-entrenamiento la escala de percepción de claves, los participantes no se vieron obligados a procesar las claves de manera sistemática antes de hacer cada juicio. Esto indica que los propios programas de entrenamiento bastan para sesgar los juicios.

CONCLUSIONES DE LA INVESTIGACIÓN REALIZADA SOBRE EL ENTRENAMIENTO EN ESTA TESIS

Al principio de este apartado se hipotetizaba que una razón que podría explicar porqué el entrenamiento a penas sí produce mejora alguna en la precisión de los juicios de veracidad, al tiempo que aumenta la probabilidad de juzgar una declaración como falsa, podría ser la propia naturaleza de los programas de entrenamiento que se ofrecen. De hecho, tanto el entrenamiento que reciben los profesionales durante su formación, como el que se ofrece en los manuales, como el que se ha venido utilizando en los experimentos de laboratorio, tienen como objetivo la detección de *mentiras*.

El objetivo de nuestros dos estudios sobre el efecto del entrenamiento era, precisamente, contrastar esta hipótesis. En este sentido, la conclusión general más importante que puede extraerse de estas dos investigaciones es que el modo en que se enfoca el entrenamiento condiciona los juicios de credibilidad.

Los programas de entrenamiento tradicionales se centran en la detección de mentiras e instruyen a los entrenandos en la identificación de claves indicadoras de mentiras. Este énfasis podría hacer que, tras el entrenamiento, los participantes buscasen activamente los indicios de engaño en los que hubieran sido entrenados, y a poco que creyeran percibirlos podrían inferir que la declaración es falsa. Aunque (en el mejor de los casos) los indicios utilizados en un entrenamiento de este tipo estén realmente relacionados con la presencia del engaño, no hay que olvidar que dicha relación es solo probabilística, y que también pueden aparecer, si bien en menor medida, cuando una persona está diciendo la verdad. De modo que la búsqueda activa de estas claves por parte de los participantes los llevaría a encontrarlas también en declaraciones verdaderas, que entonces se juzgarían como falsas.

Si la explicación anterior es correcta, entonces un programa de entrenamiento enfocado a la detección de la verdad mediante la identificación de claves de verdad debería producir un efecto opuesto; es decir, una tendencia a juzgar las declaraciones como verdaderas. Con ambos tipos de

entrenamiento, la precisión global (para verdades y mentiras conjuntamente) no tendría porque experimentar cambios significativos.

Los dos experimentos descritos en este apartado ofrecen apoyo empírico a todo el razonamiento anterior. En ambos estudios los participantes fueron entrenados para detectar mentiras (GR-TRAD) o la verdad (GR-ALT). Las claves utilizadas en los entrenamientos fueron, en realidad, los polos opuestos de cuatro indicios carentes de valor diagnóstico. A pesar de que en la fase post-entrenamiento todos los participantes vieron los mismos vídeos, realmente percibieron cosas contrarias. En concreto, los participantes de cada grupo de entrenamiento informaron haber percibido más a menudo la forma de los indicios en la que fueron entrenados (Estudio 5). Todo esto indica que los programas de entrenamiento tienen una influencia sobre las claves que los individuos entrenados perciben en una declaración durante una tarea posterior de evaluación de la credibilidad. Estas percepciones, a su vez, van a influenciar el juicio de veracidad (tal y como se ha observado en el análisis de regresión logística del Estudio 5). De hecho, después del entrenamiento los participantes del GR-TRAD hicieron más juicios de mentira que antes del entrenamiento. Por el contrario, los participantes del GR-ALT hicieron más juicios de verdad después del entrenamiento que antes de él. En el GR-CONT no se observaron cambios.

Un apoyo adicional a la hipótesis del sesgo del entrenamiento lo ofrecieron los análisis basados en la TDS (Teoría de la Detección de Señales), que mostraron que el entrenamiento tradicional sesgó significativamente los juicios de los participantes hacia la mentira, mientras que el entrenamiento alternativo los sesgó significativamente hacia la verdad.

Finalmente, el entrenamiento (independientemente de su modalidad) incrementó la confianza de los participantes en sus juicios. Por su parte, los participantes del GR-CONT, que no recibieron ningún tipo de entrenamiento, no mostraron variaciones en confianza.

En conclusión, nuestros resultados muestran que el entrenamiento puede sesgar los juicios de veracidad de los participantes e incrementar su confianza,

sin que realmente se produzca un aumento de la precisión. Esto sería especialmente perjudicial para aquellos profesionales que, por su trabajo, es probable que tengan que evaluar a menudo la credibilidad de otras personas (por ejemplo los policías). Si el entrenamiento que reciben, tanto en sus centros de formación como en los manuales, está enfocado hacia la detección de la mentira (y, como hemos visto, de hecho lo está), durante los interrogatorios y las entrevistas los policías pueden acabar tomando decisiones sesgadas (hacia la mentira) sobre la veracidad del sospechoso, y lo hacen con mucha confianza.

Estos resultados podrían interpretarse de otra manera: pudiera ser que el entrenamiento hubiera alterado las expectativas de los participantes sobre el porcentaje de declaraciones verdaderas y falsas. Por ejemplo, el fuerte énfasis del entrenamiento tradicional sobre el engaño podría haber llevado a los participantes del GR-TRAD a inferir que la mayoría de las declaraciones eran falsas. En términos más generales, y en línea con el heurístico de accesibilidad (Tversky y Kahneman, 1974), quizás el entrenamiento para detectar el engaño creó en los participantes la idea de que el engaño es muy frecuente en la vida real, mientras que el entrenamiento para detectar la verdad creó la idea de que la sinceridad es la norma. Esto pudo haber sesgado los juicios. Sin embargo, hay varios argumentos que cuestionan esta interpretación. En primer lugar, durante el entrenamiento no se ofreció ningún tipo de información sobre los porcentajes de verdades o mentiras. Por el contrario, las instrucciones que se dieron a los participantes contenían explícitamente la indicación de que cada declaración individual podría ser verdadera o falsa. En segundo lugar, los datos del Estudio 5 sugieren claramente que, al emitir sus juicios, los participantes utilizaron las claves en las que habían sido entrenados, ya que los integrantes de cada grupo percibieron más a menudo la forma de la clave en la que habían sido entrenados que la forma alternativa (Hipótesis 4), y la frecuencia percibida de las claves fue un buen predictor de los juicios (Hipótesis 5). En tercer lugar, si el entrenamiento hubiera tenido alguna influencia en las estimaciones de los participantes sobre la prevalencia de la verdad o la mentira, entonces el patrón de resultados habría sido el opuesto al encontrado. A los participantes del GR-TRAD se les indicó explícitamente al comienzo del entrenamiento que es

extremadamente difícil saber si alguien nos está mintiendo. Esto significa que la mentira es el suceso infrecuente que necesita ser detectado, lo que implica que la verdad es la norma. Esto debería haber producido un incremento en la cantidad esperada de verdades (vs. mentiras). Por el contrario, a los participantes del GR-ALT se les indicó que es extremadamente difícil saber si alguien realmente está diciendo la verdad, lo que implica la idea de que el engaño es más frecuente en la vida cotidiana de lo que tendemos a creer. Esto debería haber producido un incremento en la cantidad esperada de mentiras (vs. verdades). Por lo tanto, estas manipulaciones deberían haber llevado a los participantes del GR-TRAD a hacer más juicios de verdad y a los participantes del GR-ALT a hacer más juicios de mentira. Sin embargo, la dirección de los efectos fue la opuesta, y apoya nuestra interpretación basada en los indicios.

Al proponernos llevar a cabo una línea de investigación sobre el entrenamiento, nuestro objetivo era profundizar en el estudio del papel que Meissner y Kassin (2002) atribuían al mismo al describir el efecto del sesgo del investigador. Los resultados de los estudios que se acaban de exponer explican por qué el entrenamiento no solamente no mejora la precisión, sino que aumenta la probabilidad de juzgar una declaración como falsa.

El objetivo de esta Tesis Doctoral era la profundización en el estudio del *investigator bias effect* (efecto del sesgo del investigador), planteado por Meissner y Kassin (2002) y definido como una tendencia o sesgo de los profesionales de la detección del engaño consistente en emitir juicios de mentira al evaluar la veracidad. Meissner y Kassin (2002) demostraron que, lejos de mejorar la precisión de los juicios de veracidad, tanto el entrenamiento como la experiencia profesional en un puesto en el que sea necesario detectar mentiras no harían sino aumentar la probabilidad de que una declaración fuera juzgada como falsa.

Este trabajo ha abordado el estudio del efecto del sesgo del investigador desde una doble perspectiva. En una primera parte se ha analizado la influencia de la experiencia profesional como agente causal de este sesgo, y en segundo lugar se ha hecho lo propio con el entrenamiento.

LA EXPERIENCIA COMO POLICÍA

Se propuso, en primer lugar, una explicación basada en la sospecha. En concreto, se postuló que, de forma paralela al aumento de la experiencia profesional como policía, se iría desarrollando una sospecha comunicacional generalizada (GCS) resultante de la circunstancia de que los policías a menudo se enfrentan a situaciones donde la mentira es frecuente. Así pues, se esperaba que los policías con más años de experiencia en el cuerpo mostraran unos mayores niveles de GCS, los cuales causarían a su vez un mayor porcentaje de juicios de mentira cuando tales policías juzgaran la credibilidad.

Los resultados de la investigación realizada para contrastar esta explicación (Estudios 2 y 3) mostraron que, efectivamente, los policías con mayores niveles de experiencia como profesionales presentaban unos mayores niveles de GCS. Además, el Estudio 3 mostró que al juzgar la veracidad la precisión de los veteranos era cercana a los niveles del azar y no difería significativamente ni de la precisión de los policías noveles ni de la precisión de los estudiantes. Así

pues, en línea con el Efecto del sesgo del investigador, se puede concluir que la experiencia como policía aumenta la desconfianza de los profesionales pero no mejora su precisión como evaluadores de la veracidad.

No obstante, en el Estudio 3 también se comprobó que la GCS no determina los juicios de veracidad. Este hallazgo coincide con los del trabajo anterior de Bond y Lee (2005b).

Al observar que ambos grupos policiales emitieron un porcentaje de juicios de mentira similar y superior al de una muestra de estudiantes (Estudio 3), propusimos que es cuando se les asigna el rol social de policías que los policías noveles empiezan a comportarse tal como creen que debe hacerlo un verdadero policía, con independencia de su GCS. Es decir, es cuando realizan una tarea que perciben como “específica del trabajo policial” que los policías noveles entienden que deben comportarse como se espera de ellos en esa situación (por ejemplo, desconfiando, cuestionando y dudando de la veracidad).

Los resultados del Estudio 4 son consistentes con esta explicación, ya que muestran que, efectivamente, ante una situación “prototípicamente policial” (un interrogatorio), los policías noveles muestran elevados niveles de sospecha, independientemente de sus años de experiencia. Los resultados de este estudio explicarían a su vez lo encontrado en el Estudio 3. Probablemente, los policías noveles interpretaron la tarea de evaluación de la credibilidad del Estudio 3 como prototípica del trabajo policial (se trataba de juzgar la veracidad de una serie de declaraciones sobre un delito), por lo que “desempeñaron su papel” y se mostraron suspicaces, emitiendo tantos juicios de mentira como sus compañeros veteranos. Sin embargo, esta actitud fue una pose, ya que los policías noveles hicieron sus juicios con menor confianza que los veteranos.

EL EFECTO DEL ENTRENAMIENTO

El entrenamiento es el otro factor que Meissner y Kassin (2002) sitúan en la base del efecto del sesgo del investigador. Hemos hipotetizado que una razón que podría explicar porqué el entrenamiento tiende a sesgar los juicios de los individuos hacia la mentira podría ser la propia naturaleza de los programas de entrenamiento que se ofrecen. Se entrena a las personas para detectar

mentiras y se ofrece información sobre conductas supuestamente relacionadas con la mentira, dirigiendo así el foco atencional exclusivamente a la búsqueda tales indicios. Como resultado, al finalizar el entrenamiento las personas van a buscar activamente la presencia de aquellos indicios en los que hayan sido entrenadas (indicios de engaño, no de verdad), y van a ser especialmente sensibles a su presencia. Y a poco que crean percibirlos, emitirán probablemente un juicio de mentira. Dado que los indicios de engaño, por su naturaleza probabilística, pueden aparecer no solamente cuando alguien miente, sino también, si bien con menor frecuencia, cuando dice la verdad, las personas entrenadas también hallarán indicios de mentira en declaraciones verdaderas. Esto va a producir un incremento en los aciertos (declaraciones falsas juzgadas como tales), pero también en las falsas alarmas (declaraciones verdaderas juzgadas como falsas). Además, los cambios sobre la precisión global serán limitados.

Siguiendo este razonamiento, un programa de entrenamiento enfocado sobre la detección de la verdad (en vez del engaño) y en el que se instruyera en indicadores conductuales de verdad (en vez de indicadores de mentira), debería resultar en un aumento de la cantidad de juicios de verdad.

Para someter a contraste estos planteamientos, se realizó el Estudio 5 en el que se compararon los juicios de veracidad de dos grupos de participantes que habían sido sometidos a dos formas opuestas de entrenamiento: tradicional, enfocado en la detección de mentiras, y alternativo, enfocado hacia la detección de la verdad. Se halló que el entrenamiento tradicional incrementaba el porcentaje de juicios de mentira, mientras que el alternativo lo reducía.

Un segundo estudio (Estudio 6) realizado siguiendo los mismos parámetros que en el anterior ofreció una réplica de los resultados previos, aun cuando en este caso se había modificado el formato de los cuestionarios de evaluación post-entrenamiento para tener la certeza de que ningún factor salvo el tipo de entrenamiento era responsable de las variaciones en los juicios de veracidad.

En definitiva, la investigación sobre el entrenamiento presentada aquí explica porqué los programas de entrenamiento tradicionales, centrados sobre

la detección de mentiras, generan, en último término, un sesgo hacia la mentira y no un incremento de la precisión.

NECESIDAD DE SUPERAR EL EFECTO DEL SESGO DEL INVESTIGADOR

Superar, o al menos tratar de minimizar el efecto del sesgo del investigador implicaría reducir el sesgo de los profesionales al evaluar la credibilidad y aumentar la precisión de sus juicios. Esto significaría un mejor desempeño de su labor, lo que a su vez se traduciría en múltiples ventajas para todas las personas que nos beneficiamos de su trabajo. Haber observado cómo actúan las variables que la investigación ha situado en la base de este sesgo (la experiencia y el entrenamiento), ofrece a su vez pistas respecto a cómo enfrentarse a él.

RESPECTO A LA EXPERIENCIA

La experiencia profesional es un valor muy tenido en cuenta dentro de la institución policial. Para los nuevos reclutas, los compañeros veteranos suponen a menudo un complemento a la instrucción formal. Por ejemplo, los datos de Colwell, Miller, Lyons, et al. (2006) mostraban que muchos policías que no habían recibido un entrenamiento formal en detección del engaño indicaron haber sido instruidos por sus compañeros más veteranos. De esta forma, los policías con más experiencia pueden constituir un modelo a la hora de gestionar situaciones que, en los centros de formación, sólo se han abordado de forma teórica o con prácticas simuladas. En este sentido, ya hemos visto anteriormente que los policías a penas reciben instrucción sobre la evaluación de la credibilidad. No obstante, la investigación ha mostrado que la experiencia policial no implica una mayor precisión al juzgar la veracidad de una declaración, y sí un aumento de la GCS. Los policías más veteranos podrían entonces contribuir al progresivo desarrollo de una GCS en los policías más noveles. ¿Cómo abordar este problema? ¿Cómo transmitir a los policías noveles que las enseñanzas y ejemplos de sus compañeros más veteranos pueden perjudicarles? Esta es, sin duda, una tarea difícil.

Una forma de afrontar el problema sería informar a los profesionales sobre sus propios sesgos, mostrándoles la evidencia sobre el papel de la experiencia profesional en el desarrollo de la GCS, y demostrándoles que ser más veterano no implica ser mejor evaluador de la credibilidad. No obstante, probablemente no bastará con presentar la evidencia científica disponible⁵⁹. En este sentido, una medida eficaz podría ser someter a los propios profesionales a un proceso experimental similar al del Estudio 3, en el que policías noveles y veteranos completaron un cuestionario de GCS y posteriormente juzgaron la veracidad de distintas declaraciones. De este modo advertirán por sí mismos sus limitaciones. Enfrentar a los profesionales a los resultados de estudios en los que ellos mismos han sido los participantes y comprobar que coinciden con los de estudios previos podría ser más eficaz que, simplemente, presentarles los resultados de otras investigaciones.

RESPECTO AL ENTRENAMIENTO

La evidencia científica sobre el efecto del entrenamiento apunta indudablemente a la necesidad de redefinir y reestructurar el modo en que se entrena a las personas a diferenciar entre verdades y mentiras. Los programas tradicionales, en los que simplemente se ofrece información sobre la relación de determinadas conductas con el engaño, deben ser sustituidos por un nuevo concepto de entrenamiento en el que el primer elemento a redefinir es la propia etiqueta “detección del engaño”. A continuación presentamos algunos aspectos a tener en cuenta para abordar esta difícil tarea.

a) En primer lugar, antes instruir a los profesionales sobre cómo es más probable que se comporte una persona al mentir y al decir la verdad, es necesario comenzar eliminando las falsas creencias y estereotipos que tienen estos profesionales. Mientras estas falsas creencias existan y, tal como Strömwall et al. (2004) indican, se sigan transmitiendo de policías veteranos a

⁵⁹ A lo largo de la realización de esta Tesis y en las diversas ocasiones anteriores en las que algún miembro de nuestro equipo de investigación se ha desplazado al Centro de Formación de la Policía de Ávila, se ha observado una reticencia en los policías a aceptar los resultados de la investigación científica cuando contradicen sus propias creencias, experiencias y expectativas. Asimismo, ya se ha indicado anteriormente que Vrij (2000), que ha hecho numerosas investigaciones con policías, advierte que éstos pueden no creer los datos que les ofrece el instructor universitario.

policías noveles, es muy probable que tales policías sencillamente no confíen en el entrenamiento (Vrij, 2000), especialmente si éste cuestiona las enseñanzas de los compañeros más veteranos.

b) Una vez que se haya tratado el problema de las falsas creencias, es cuando realmente debe comenzar a administrarse el entrenamiento en sí mismo. Distintos investigadores han definido ya las características que debe reunir un buen programa de entrenamiento para detectar mentiras (véase, por ejemplo, Frank y Feeley, 2003). Aunque lo que defendemos aquí es que no debe entrenarse solamente en la detección del engaño, sino también en la identificación de la verdad, las recomendaciones más importantes que Frank y Feeley (2003) plantean (ver Anexo R) pueden y deben aplicarse a cualquier programa de entrenamiento.

La idea principal que aportamos desde estas páginas a la hora de construir un programa de entrenamiento es que un protocolo que incluyera la instrucción en conductas indicadoras de verdad compensaría el sesgo hacia la mentira característico de los entrenamientos tradicionales. Esto haría que, al enfrentarse a una tarea de evaluación de la veracidad, las personas buscaran activamente no sólo indicios de engaño, sino también de verdad. Por otro lado, la investigación ha mostrado que los policías tienden a buscar indicios de engaño y no de veracidad, e interpretan indicios ambiguos como signos de mentira (Masip, Garrido, Herrero, Antón, et al. 2006). Si se les entrenara no sólo en indicios de mentira sino también en indicios de verdad, entonces probablemente emitirían sus juicios basándose en una búsqueda y comparación de ambos tipos de indicios. Así, cualquiera que fuera el juicio final emitido, no estaría basado en una percepción “mínima” de un solo tipo de conducta. Es decir, si bien es probable que un policía tras un entrenamiento tradicional juzgue una declaración como falsa a partir de la observación de un solo indicio de engaño en el que ha sido entrenado, tras un entrenamiento que incluyera los dos tipos de indicios, sería probable que no infiriera engaño si además de dicho indicio de mentira percibiera también un indicio de verdad.

c) No obstante todo lo dicho, uno de los problemas más importantes a los que ha de enfrentarse un programa de entrenamiento es la ausencia de

indicios absolutamente fiables de mentira o de verdad. De hecho, como se ha visto, incluso los indicios que han mostrado mayor poder discriminativo tienen una utilidad solo moderada. A menudo, los meta-analistas utilizan el índice d propuesto por Cohen (1988)⁶⁰ como una medida del tamaño del efecto. En el meta-análisis de DePaulo, Lindsay, et al. (2003) solamente dos indicios mostraron un tamaño de efecto superior a $|0.50|$ (tamaño indicado por Cohen para poder hablar de un efecto grande). Los recientes meta-análisis de Sporer y Schwandt (2006, 2007), también mostraron que la asociación de los indicios observables con el engaño es pequeña. El valor medio de d fue $|0.097|$ para los indicios paralingüísticos y $|0.089|$ para los visibles. Richard, Bond y Stokes-Zoota (2003) examinaron 322 meta-análisis en psicología social y establecieron una d de 0.20 como pequeña, de 0.40 como media y de 0.60 como grande, ya que esos valores eran mayores que el valor promedio de d encontrado en el 30%, 50% y 75% de los meta-análisis que examinaron. A la vista de esto, los tamaños del efecto medios encontrados por DePaulo, Lindsay, et al. (2003) y Sporer y Schwandt (2006, 2007) son ciertamente pequeños. Esto supone un serio problema: si no hay indicadores válidos de engaño, ¿qué indicios deberían utilizarse para los programas de entrenamiento? La ausencia de indicadores válidos a nivel general es un problema que quizá podría solucionarse si los programas de entrenamiento incluyeran información sobre variables moderadoras, ya que en todos los meta-análisis recientes sobre las claves del engaño las pruebas de heterogeneidad han sido significativas, y los análisis sobre los moderadores han mostrado que hay indicios que discriminan en ciertas circunstancias y no en otras (véanse DePaulo, Lindsay, et al., 2003; Sporer y Schwandt, 2006, 2007).

Además de instruir a los entrenandos sobre las situaciones particulares en las que discrimina cada clave, debería enseñárseles cómo crear situaciones en las cuales van a aparecer ciertos indicios de engaño. En este sentido, los trabajos de Hartwig et al. (2006) y Vrij et al. (2008) son particularmente relevantes. Hartwig et al. (2006) se centran en las circunstancias en las cuales aparecen inconsistencias en los mentirosos y éstas desvelan el engaño. Según

⁶⁰ En general se trata de la media estandarizada de las diferencias entre el grupo que ha recibido el tratamiento y el grupo control. En nuestro caso particular es la media estandarizada de las diferencias en el grado de presencia de una conducta al mentir y decir la verdad.

los autores, en lugar de acusar directamente al sospechoso que está siendo interrogado para que éste confiese, la policía no debería desvelar la evidencia incriminatoria disponible hasta el final del interrogatorio. Esto haría que el sospechoso no supiera que la policía tiene ciertas pruebas contra él, y que sus declaraciones (por ejemplo, que el día del delito viajó con su automóvil fuera de la ciudad) contradigan lo que la policía ya sabe (por ejemplo, que su automóvil fue visto el día del delito cerca del lugar de los hechos).

Hartwig et al. (2006) sometieron a contraste empírico estas ideas y encontraron que, en comparación con un grupo de policías no entrenados, los policías que habían sido entrenados para no desvelar la evidencia hasta el final de la entrevista lograron en mayor medida no hacerlo cuando entrevistaban a mentirosos que cuando entrevistaban a personas sinceras. Además, también lograron generar más inconsistencias en los mentirosos (pero no en personas veraces, cuyas declaraciones fueron siempre bastante consistentes con los hechos) que los policías no entrenados. Así mismo, a diferencia del grupo no entrenado, los entrevistadores entrenados hicieron más juicios de mentira cuantas más inconsistencias mostraron los entrevistados. Finalmente, al juzgar verdades y mentiras, los entrevistadores entrenados alcanzaron índices de precisión (85.4%) sensiblemente superiores que los de los entrevistadores no entrenados (56.1%). Estos resultados son prometedores, ya que muestran que, bajo ciertas circunstancias, el entrenamiento (pero no el clásico tipo de entrenamiento que se ha empleado habitualmente) puede ayudar a los profesionales a detectar mentiras.

El Trabajo de Vrij et al. (2008) va en la misma línea, al mostrar que si los emisores hacen las declaraciones en orden inverso respecto a la cronología de los hechos, entonces las diferencias conductuales entre mentirosos y veraces aumentan. Cuando tales declaraciones “inversas” se mostraron a una muestra de policías para que evaluaran su credibilidad, sus niveles de precisión fueron superiores que los de otra muestra de policías que tuvo que juzgar la veracidad de las declaraciones en orden cronológico. El proceso que subyace a estos efectos es la carga cognitiva, la cual da lugar a la aparición de ciertos indicios conductuales. Al pedir a los declarantes que cuenten sus historias en orden inverso, se incrementa la carga cognitiva asociada a la tarea. Dado que mentir

es ya en sí mismo una labor que demanda más recursos cognitivos que decir la verdad, la sobrecarga cognitiva se hace especialmente notable en aquellos casos en los que la historia es falsa, lo que se traduce en un aumento de las diferencias conductuales entre quienes mienten y quienes dicen la verdad. Esto, a su vez, facilita a los evaluadores la discriminación entre verdades y mentiras.

Los programas de entrenamiento deberían tener en cuenta los hallazgos científicos (como el efecto de aumentar la carga cognitiva sobre la conducta de quienes tienen que emitir una declaración) e incluir estrategias como las empleadas por Hartwig et al. (2006) y Vrij et al. (2008) destinadas a favorecer la aparición de indicios útiles para discriminar entre verdades y mentiras.

Otra solución al problema de la escasa discriminabilidad de las claves parte de los resultados de investigaciones recientes que muestran que aunque los indicios conductuales son, a nivel individual, pobres predictores de la veracidad de una declaración, determinadas configuraciones o conjuntos de indicios podrían tener una mayor utilidad predictiva (véase Rothwell, et al., 2006). Quizá los programas de entrenamiento deberían centrarse en el poder discriminativo de varios indicios tomados conjuntamente en vez de instruir a los profesionales en indicios aislados. No obstante, en el trabajo de Rothwell et al. (2006) fue un ordenador el que, a través de una red neuronal artificial, estableció y trabajó con configuraciones de indicios. Quizá ésta sea una tarea demasiado compleja para los seres humanos. En cualquier caso, la investigación sobre el poder discriminativo de configuraciones de indicios está aún empezando a desarrollarse, y es temprano para establecer conclusiones definitivas sobre qué configuraciones o conjuntos de conductas son más eficaces y la posibilidad de que los seres humanos puedan ser entrenados con ellas.

d) El estudio anteriormente descrito de Hartwig et al. (2006) enlaza con otra cuestión importante: los programas de entrenamiento deberían diseñarse teniendo en cuenta que para hacer un juicio de veracidad los profesionales no se apoyan solamente en la forma y el contenido de las declaraciones, sino que, en ocasiones, cuentan con otros recursos, como pruebas físicas que sitúan o no al interrogado en el lugar de los hechos, grabaciones de cámaras de

seguridad o testimonios de terceras personas. En este sentido, los resultados del estudio de Park, Levine, McCornack, Morrison y Ferrara (2002) son particularmente reveladores. Los autores trataron de examinar el modo en que las personas “realmente” detectan el engaño. Para ello pidieron a 202 participantes que recordasen una situación cotidiana pasada en la que hubieran descubierto que otra persona les había mentado y que indicasen cómo lo habían descubierto. Los datos mostraron que la información de terceras personas, la evidencia física disponible o la combinación de ambas eran modos mucho más habituales de descubrir una mentira que la mera observación de la conducta verbal y no verbal del mentiroso durante el episodio de engaño.

Al igual que los participantes del estudio de Park et al. (2002), los policías que tratan de dilucidar si un testimonio es verdadero o falso dispondrán de otras fuentes de información que pueden ser determinantes a la hora de emitir un juicio de veracidad. Un buen programa de entrenamiento debe considerar esta circunstancia e instruir a los profesionales en un manejo adecuado de la evidencia adicional fruto de la investigación policial. En este sentido, el programa de entrenamiento de Hartwig et al. (2006) que hemos descrito unas páginas más atrás es ejemplar.

CONCLUSIONES

Pensamos que hay una serie de consideraciones que pueden contribuir a la superación, o al menos la reducción, del efecto del sesgo del investigador:

Respecto a la experiencia profesional, se puede tratar de minimizar su efecto pernicioso tratando que los propios profesionales tomen conciencia de sus efectos negativos para la evaluación de la credibilidad. Esta es, sin duda, una tarea muy difícil dadas las características de este grupo y el valor y credibilidad que tradicionalmente se otorga a los compañeros más veteranos. No obstante, el mostrar empíricamente a los policías su falibilidad es un posible camino a seguir.

Respecto al entrenamiento, se plantea la necesidad de una redefinición del mismo, teniendo en cuenta que (a) debe comenzarse eliminando las falsas creencias, especialmente sobre las conductas que indican engaño; (b) deben

diseñarse protocolos para la evaluación de la veracidad, es decir, hay que entrenar a los profesionales para identificar tanto mentiras como verdades, y no solamente para detectar el engaño; (c) ante la falta de indicios conductuales totalmente fiables, el entrenamiento debe tener en cuenta los hallazgos de la investigación científica, especialmente aquellos referidos a qué conductas tienen un mayor valor diagnóstico, en qué situaciones, y si existen combinaciones de indicios cuya utilidad predictiva sea superior a la de éstos a nivel individual; (d) debe instruirse a los policías en el empleo de aquellas estrategias que la investigación ha mostrado que son de utilidad, tanto en un primer momento a la hora de interrogar a un sospechoso (para hacer aflorar indicios conductuales útiles para discriminar entre verdades y mentiras) como posteriormente a la hora de tomar una decisión sobre la veracidad de la declaración; y (e) el entrenamiento debe incluir la instrucción en la utilización de los recursos específicos con los que, por las características de su trabajo, los policías puedan contar, por ejemplo, la evidencia obtenida a partir de la investigación policial del caso.

El objetivo último que persiguen los requisitos que se acaban de proponer, y en general toda esta Tesis Doctoral, es afrontar un importante obstáculo para una labor policial más eficaz: el efecto del sesgo del investigador. Este obstáculo se hace particularmente patente en aquellas circunstancias en las que los profesionales deben tomar una decisión sobre la veracidad de una declaración. Es precisamente en tales circunstancias en las que una decisión errónea puede perjudicar a un inocente, o hacer que un culpable quede indemne. Para tratar de evitar que esto ocurra, ha de darse una extraña circunstancia absolutamente imprescindible: que científicos sociales y agentes de la ley trabajen juntos.

BIBLIOGRAFÍA

- Aamodt, M. G. y Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner*, 15, 6-11.
- Abrunhosa, R., Soeiro, C. y Ventura, J. P. (2006). Psicología de la policía. En E. Garrido, J. Masip y C. Herrero (Eds.), *Psicología jurídica* (pp. 111-129). Madrid: Pearson.
- Adams, S. H. (1996). Statement analysis: What do suspects' words really reveal? *FBI Law Enforcement Bulletin*, Octubre, 12-20.
- Akehurst, L., Bull, R., Vrij, A. y Köhnken, G. (2004). The effects of training professional groups and lay persons to use criteria-based content analysis to detect deception. *Applied Cognitive Psychology*, 18, 877-891.
- Alonso-Quecuty, M. L. (1994). Psicología forense experimental: El testigo deshonesto. En J. Sobral, R. Arce y A. Prieto (Eds.), *Manual de psicología jurídica* (pp. 139-153). Barcelona: Paidós.
- Alvarez, M. (2000, Julio). *Psicología y seguridad pública*. Trabajo presentado en el I Congreso Hispano Alemán de Psicología Jurídica, Pamplona, Navarra.
- Amnesty International (2000). *Torture worldwide: An affront to human dignity*. Londres: Autor.
- Amnistía Internacional (2008). *El estado de los derechos humanos en el mundo*. Madrid: Autor.
- Anderson, M. (1986). Cultural concatenation of deceit and secrecy. En R. W. Mitchell y N. Thomsom (Eds.), *Deception, perspectives on human and non human deceit* (pp. 323-348). Albany: State University of New York Press.
- Anderson, J., Kuroshima, H., Kuwahata, H. y Figita, K. (2001). Training squirrel monkeys (*saimiri sciureus*) to deceive: Acquisition and analysis of behavior toward cooperative and competitive trainers. *Journal of Comparative Psychology*, 115, 282-293.
- Barnier, A. J., Sharman, S. J., McKay, L. y Sporer, S. L. (2005). Discriminating adults' genuine, imagined, and deceptive accounts of positive and negative childhood events. *Applied Cognitive Psychology*, 19, 985-1001.
- Beals, R. L. y Hoijer, H. (1978). *Introducción a la antropología*. Madrid: Aguilar.
- Bill Clinton admite que mintió. (2001, 20 de Enero). *Prensa Web*. Extraído el 20 de Mayo del 2004 de <http://mensual.prensa.com/mensual/contenido/2001/01/20/hoy/portada>

- Blair, J. 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. y DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214-234.
- Bond, C. F., Kahler, K. N. y Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology*, 21, 331-345.
- Bond, C. F., Omar, A., Mahmoud, A. y Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior*, 14, 189-204.
- Bond, C. F. y Uysal, A. (2007). On lie detection "wizards". *Law and Human Behavior*, 31, 109-115.
- Bond, G. D. (2007). Deception detection expertise. *Law and Human Behavior*, 32, 339-351.
- Bond, G. D. y Lee, A. Y. (2005a). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, 19, 313-329.
- Bond, G. D. y Lee, A. Y. (2005b). The darkest side of trust: Validating the generalized communication suspicion scale with prison inmates. *Personality and Individual Differences*, 38, 1429-1438.
- Bond, G. D., Malloy, D., Arias, E., Nunn, S. y Thompson, L. (2005). Lie-biased decision making in prison. *Communication Reports*, 18, 9-19.
- Bull, R. (1989). Can training enhance the detection of deception? En J. C. Yuille (Ed.), *Credibility assessment* (pp. 83-99). Dordrecht: Kluwer Academic Publishers.
- Bull, R. (1999). Police investigative interviewing. En A. Memon, y R. Bull (Eds.), *Handbook of the psychology of interviewing* (pp. 279-291). Chichester, Reino Unido: John Wiley and Sons.
- 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 polygraphic deception detection*. Londres: British Psychological Society.

- Buller, D. B. y Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. En J. A. Daly y J. M. Wiemann (Eds.), *Strategic interpersonal communication* (pp. 191-223). Hillsdale, NJ: Erlbaum.
- Bush, G., Whalen, P., Rosen, B., Jenike, M., McInerney, S. y Rauch, S. (1998). The counting stroop: An interference task specialized for functional neuro-imaging – validation study with functional fMRI. *Human Brain Mapp*, 6, 270-282.
- Carrancho, J. I. (2002). La entrevista policial. En *Psicología Social II. Segundo curso Escala Ejecutiva*. Ávila: División de Formación y Perfeccionamiento de la D.G.P.
- Centro de promoción. La carrera policial. (sin fechar). Extraído el 3 de Mayo del 2008 de <http://www.policia.es/dfp/index.htm>
- Chevalier-Skolnikoff, S. (1986). An exploration of the ontogeny of deception in human beings and nonhuman primates. En R. W. Mitchell y N. Thomson (Eds.), *Deception, perspectives on human and nonhuman deceit* (pp. 205-220). Albany, NY: State University of New York Press.
- Christie, R. y Geis, F. (1970). *Studies in machiavellianism*. Nueva York: Academic Press.
- Clemente, M. (1995). *Fundamentos de la psicología jurídica*. Madrid: Pirámide.
- Clinton: La hora de la verdad. (1998, 21 de Septiembre). *QuePasa*. Extraído el 15 de Mayo del 2004 de <http://www.quepasa.cl/revista/1432/9.html>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Ed. rev.). Hillsdale, NJ: Erlbaum.
- Cole, T. (2001). Lying to the one you love: The use of deception in romantic relationships. *Journal of Social and Personal Relationships*, 18, 107-129.
- Coleman, L. y Kay, P. (1981). Prototype semantics: The English word lie. *Language*, 57, 26-44.
- Colwell, L., Miller, H., Lyons, P. y Miller, R. (2006). The training of law enforcement officers in detecting deception: A survey of current practices and suggestions for improving accuracy. *Police Quarterly*, 9, 275-290.
- Colwell, L., Miller, H., Miller, R. y Lyons, P. (2006). US police officers' knowledge regarding behaviors indicative of deception: Implications for eradicating erroneous beliefs through training. *Psychology, Crime and Law*, 12, 489-503.

- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughhead, R. C., et al. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage*, 28, 663-668.
- De Lumley, H. (2000). *El primer hombre*. Madrid: Cátedra.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111, 203-243.
- DePaulo, B. M. (2004). The many faces of lies. En A. G. Miller (Ed.), *The social psychology of good and evil: Understanding our capacity for kindness* (pp. 303-326). Guilford.
- 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, 346-357.
- DePaulo, B. M. y Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74, 63-79.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S., Wyer, M. y Epstein, J. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979-995.
- DePaulo, B. M., Lassiter, G. D. Y Stone, J. I. (1982). Attentional determinants of success at detecting deception and truth. *Personality and Social Psychology Bulletin*, 8, 273-279.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. y Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- DePaulo, B. M. y Morris, W. L. (2004). Discerning lies from truths: Behavioral cues to deception and the indirect pathway of intuition. 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. En B. R. Schlenker (Ed.), *The self and social life* (pp. 323-370). McGraw-Hill.
- DePaulo, B. M., Wetzel, C., Weylin R. y Walter J. (2003). Verbal and nonverbal dynamics of privacy, secrecy and deceit. *Journal of Social Issues*. 59, 391-410.

- DePaulo, B. M., Zuckerman, M. y Rosenthal, R. (1980). Humans as lie detectors. *Journal of Communication*, 30, 129-139.
- División de Formación y Perfeccionamiento de la D.G.P. (2000). *Diseño curricular para la formación de la Escala Ejecutiva del Cuerpo Nacional de Policía. Y Plan de Estudios 2006*. Ávila: Centro de Formación de la Policía.
- División de Formación y Perfeccionamiento de la D.G.P. (2006). *Plan de estudios 2000, Escala Ejecutiva*. Ávila: Centro de Formación de la Policía.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, 121, 275-277.
- Doner, J. (1998). Are we lying to ourselves about deception? *Social Service Review*, 72, 234-250.
- Driscoll, L. N. (1994). A validity assessment of written statements from suspects in criminal investigations using the SCAN technique. *Police Studies*, 17, 77-88.
- Ekman, P. (1992). *Cómo detectar mentiras. Una guía práctica para utilizar en el trabajo, la política y la pareja* (L. Wolfson, Trad.). Barcelona: Paidós. (Trabajo original publicado en 1985).
- Ekman, P. (1997a). Deception, lying, and demeanor. En D. F. Halpern, y A. E. Voiskounsky (Eds.), *States of mind. American and post-Soviet perspectives on contemporary issues in psychology* (pp. 93-105). Nueva York: Oxford University Press.
- Ekman, P. (1997b). Lying and deception. En N. L. Stein, P. A. Ornstein, B. Tversky y C. Brainerd (Eds.), *Memory for everyday and emotional events*. Mahwah, NJ: Erlbaum.
- Ekman, P. y Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29, 288-298.
- Ekman, P. y O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913-920.
- Ekman, P., O'Sullivan, M. y Frank, M. (1999). A few can catch a liar. *Psychological Science*, 10, 263-266.
- Elaad, E. (2003). Effects of feedback on the overestimated capacity to detect lies and the underestimated ability to tell lies. *Applied Cognitive Psychology*, 17, 349-363.

- Elliot, R., Dolan, R. y Frith, C. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, 10, 308-317.
- Evans, C. y Harwood, J. (Eds.). (1991). *The Collins Spanish dictionary*. Barcelona: Grijalbo.
- Farwell, L. y Smith, S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Sciences*, 46, 135-143.
- Feeley, T. H. y DeTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Communication Quarterly*, 43, 420-430.
- Feeley, T. H. y Young, M. (2000). Self-reported cues about deceptive and truthful communication: The effects of cognitive capacity and communicator veracity. *Communication Quarterly*, 18, 101-119.
- Fernández, M. (2003). La investigación policial del delito: Competencias de la policía en la investigación del delito y métodos policiales de investigación. En J. Burón (Ed.), *Psicología médico-forense. La investigación del delito* (pp. 285-347). Bilbao: Desclée de Brouwer.
- Fernández, M. (1990). *La profesión de policía*. Madrid: CIS.
- Fiedler, K. y Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, 20, 199-223.
- Ford, E. B. (2006). Lie detection: Historical, neuropsychiatric and legal dimensions. *International Journal of Law and Psychiatry*, 29, 159-177.
- 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, 58-75.
- Ganis, G., Kosslyn, S., Stose, S., Thompson, W. y Yurgelun-Todd, D. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex*, 13, 830-836.
- Garrido, E. y Herrero, C. (2006). Relaciones entre la psicología y la ley. En E. Garrido, J. Masip y C. Herrero (Eds.), *Psicología jurídica* (pp. 4-42). Madrid: Pearson.
- Garrido, E. y Masip, J. (1999). How good are police officers at spotting lies? *Forensic Update*, 58, 14-21.

- Garrido, E. y Masip, J. (2001). La evaluación psicológica en los supuestos de abusos sexuales. En F. Jiménez (Ed.), *Evaluación psicológica forense 1: Fuentes de información, abusos sexuales, testimonio, peligrosidad y reincidencia* (pp. 25-140). Salamanca: Amarú.
- Garrido, E., Masip, J., Herrero, C. y Rojas, M. (2000). La detección del engaño a partir de claves conductuales por agentes de policía. En A. Ovejero, M. V. Moral y P. Vivas (Eds.), *Aplicaciones en psicología social* (pp. 97-105). Madrid: Biblioteca Nueva.
- Garrido, E., Masip, J. y Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology*, 39, 254-275.
- Geary, J. (2000, 13 de Marzo). Deceitful minds, the awful truth about lying. *Magazine Europe*. Extraído el 9 de Diciembre del 2002 de <http://www.time.com/time/europe/magazine/2000/313/deceit.html>
- George, J. F., Marett, K., Burgoon, J. K., Crews, J., Cao, J., Lin, M., et al. (2004). Training to detect deception: An experimental investigation. *Proceedings of the 37th Hawaii International Conference on System Sciences*. Extraído el 26 de Noviembre de 2008 de <http://www2.computer.org/plugins/dl/pdf/proceedings/hicss/2004/2056/01/205610023a.pdf?template=1&loginState=1&userData=anonymous-IP1227716299087>.
- Gispert, C., Gay, J. y Vidal, J. (Eds.). (1996). *Diccionario de sinónimos y antónimos*. Barcelona: Océano.
- Goffman, E. (1959). *The presentation of self in everyday life*. Nueva York: Anchor Books.
- Granhag, P., Anderson, L., Strömwall, L. y Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminal Psychology*, 9, 1-17.
- Granhag, P. A. y Strömwall, L. A. (2004). Deception – a central characteristic of human life. En P. A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 3-12). Cambridge: Cambridge University Press.
- Granhag, P. A. y Vrij, A. (2005). Deception detection. En N. Brewer y K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 43-92). Nueva York: Guilford.
- Green, D. M. y Swets, J. A. (1966). *Signal detection theory and psychophysics*. Nueva York: John Wiley and Sons.

- Grigorieff, V. (1995). *El gran libro de las religiones del mundo* (T. Clavel, Trad.). Barcelona: Robin Book. (Trabajo original publicado en 1989).
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions*. Chichester, Reino Unido: John Wiley and Sons.
- Hancock, J. T., Curry, L., Goorha, S. y Woodworth, M. T. (2004). Lies in conversation: An examination of deception using automated linguistic analysis. En K. Forbus, D. Gentner y T. Regier (Eds.), *Proceedings Mahwah, NJ: Erlbaum*.
- Haney, C., Banks, W. C. y Zimbardo, P. G. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1, 69-97.
- Hartwig, M., Granhag, P. A., Strömwall, L. F. y Kronkvist, O. (2006). Strategic use of evidence during police Interviews: When training to detect deception works. *Law and Human Behavior*, 30, 603-619.
- Hartwig, M., Granhag, P. A., Strömwall, L. A. y Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, 7, 429-456.
- Hedges, L. y Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Honts, C. R. (2004). The psychophysiological detection of deception. En P. A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 103-123). Cambridge: Cambridge University Press.
- Hopper, R. y Bell, R. A. (1984). Broadening the deception construct. *Quarterly Journal of Speech*, 70, 288-302.
- Horowitz, S. W. (1991). Empirical support for statement validity assessment. *Behavioral Assessment*, 13, 293-313.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuck, T., Krispin, O. y Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, 2, 11-21.
- Horvath, F., Jayne, B. y Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavioral analysis interviews. *Journal of Forensic Sciences*, 39, 793-807.
- Inbau, F. E., Reid, J. E. y Buckley, J. P. (1986). *Criminal interrogation and confessions* (3ª ed.). Baltimore, MD: Williams and Wilkins.

- Inbau, F. E., Reid, J. E., Buckley, J. P. y Jayne, B. C. (2001). *Criminal interrogation and confessions* (4ª ed.). Gaithersberg, MD: Aspen.
- Jensen, M. L., Meservy, T. O., Kruse, J., Burgoon, J. K. y Nunamaker, J. F. (2005). Lessons learned from developing systems to automatically detect deception in various modalities. *The Business Review*, 4, 37-43.
- John E. Reid and Associates. (sin fechar). *Intreviewing and interrogation*. Extraído el 11 de abril del 2008 de http://www.reid.com/training_programs/interview_overview.html
- Johnson, M. K. y Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Kalbfleisch, P. J. (1985). Accuracy in deception detection: A quantitative review. *Dissertation Abstracts International*, 45, 46112B. (UMI No. 8603433).
- Kashy, D. A. y DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology*, 70, 1037-1051.
- Kassin, S. M. (2002). Human judges of truth, deception, and credibility: Confident but erroneous. *Cardezo Law Review*, 23, 809-816.
- Kassin, S. M. y College, W. (2005). On the psychology of confessions. Does innocence put innocents at risk? *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., Goldstein, C. C. y Savitsky, K. (2002). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, 27, 187-203.
- Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A. M., et al. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior*, 31, 381-400.
- Kassin, S. M., Meissner, C. y Norwick, R. (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.
- Kerr, P. (1990). *Book of lies*. Londres: Penguin.

- Kircher, J. C. y Raskin, D. C. (2002). Computer methods for the psychophysiological detection of deception. En M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 287-326). Londres: Academic Press.
- Kirsch, J. P. (1911). Ordeals. En *The Catholic Encyclopedia* (vol. XI). Robert Appleton Company. Extraído el 28 de Septiembre del 2003, de <http://www.newadvent.org/cathen/11276b.htm>
- Knapp, M. L., Hart, R. P. y Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Research*, 1, 15-29.
- Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, 2, 1-17.
- Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms, and results. En H. Wegener, F. Lösel y J. Haisch (Eds.), *Criminal behavior and the justice system* (pp. 271-289). Londres: Springer-Verlag.
- Köhnken, G. (2004). Statement validity analysis and the detection of the truth. En P. A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41-63). Cambridge: Cambridge University Press.
- 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.
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J. y George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58, 605-613.
- Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, 30, 209-216.
- La aplicación del LIWC. (sin fechar). Extraído el 19 de Noviembre del 2007 de <http://www.liwc.net/liwcspanol/liwcdescription.php>
- Lamb, M. E., Sternbeg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y. y Hovav, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect*, 21, 225-264.
- Langbein, J. H. (1976). *Torture and the law of proof*. Chicago, IL: The University of Chicago Press.
- Langleben, D., Schroeder, L., Maldjian, J., Gur, R., McDonald, S., Ragland, J., et al. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *Neuroimage*, 15, 727-732.

- Larson, J. A. (1969). *Lying and its detection. A study of deception and deception tests*. Montclair, NJ: Patterson Smith. (Trabajo publicado originalmente en 1932 por The University of Chicago Press, Chicago).
- Larzelere, R. E. y Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, 42, 595-604.
- Leach, A., Talwar, V., Lee, K., Bala, N. y Lindsay, R. (2004). "Intuitive" lie detection of children's deception by law enforcement officials and university students. *Law and Human Behavior*, 28, 661-685.
- Lee, T. M., Liu, H. L., Tan, L. H., Chan, C. C., Mahankali, S., Feng, C. M., et al. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping*, 15, 157-164.
- Lesce, A. (1990). SCAN: Deception detection by scientific content analysis. *Law and Order*, 38. Extraído el 7 de Noviembre del 2007 de <http://www.lscan.com/id37.htm>
- Levine, T. R. y McCornack, S. A. (1991). The dark side of trust: Conceptualizing and measuring types of communicative suspicion. *Communication Quarterly*, 39, 325-339.
- 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.
- Lykken, D. T. (1998). *A tremor in the blood. Uses and abuses of the lie detector*. Nueva York: Plenum Press.
- MacDonald, A., Cohen, J., Stenger V. y Carter, C. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288, 1853-1858.
- Mann, S. y Vrij, A. (2006). Police officers' judgements of veracity, tenseness, cognitive load and attempted behavioral control in real-life police interviews. *Psychology, Crime and Law*, 12, 307-319.
- 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.
- Mann, S., Vrij, A., Fisher, P. y Robinson, M. (2007). See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. *Applied Cognitive Psychology*, 22, 1062-1071.

- Manzanero, A. (2001). Procedimientos de evaluación de la credibilidad de las declaraciones de menores víctimas de agresiones sexuales. *Psicopatología Clínica, Legal y Forense*, 1, 51-71.
- Martín-Albo, M. (2003). *La masonería, una hermandad de carácter secreto*. Madrid: LIBSA.
- Masip, J. (2002). *La evaluación de la credibilidad del testimonio a partir de los índices conductuales en el contexto jurídico penal*. Tesis doctoral no publicada. Universidad de Salamanca.
- Masip, J. (2005). ¿Se pilla antes a un mentiroso que a un cojo? Sabiduría popular frente a conocimiento científico sobre la detección no-verbal del engaño. *Papeles del Psicólogo*, 26, 78-91.
- Masip, J. y Alonso, H. (2006). Verdades mentiras y su detección: Aproximaciones verbales y psicofisiológicas. En E. Garrido, J. Masip y C. Herrero (Eds.), *Psicología jurídica* (pp. 505-558). Madrid: Pearson.
- Masip, J., Alonso, H., Garrido, E. y Antón, C. (2005). Generalised communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, 25, 1046-1066.
- Masip, J., Alonso, H. y Herrero, C. (2006). Verdades, mentiras y su detección a partir del comportamiento no verbal. En E. Garrido, J. Masip y C. Herrero (Eds.), *Psicología jurídica* (pp. 475-505). Madrid: Pearson.
- Masip, J. y Garrido, E. (1999). Evaluación psicológica forense de la credibilidad: Contextualización teórica y paradigmas evaluativos. En A. P. Soares, S. Araujo y S. Caires (Eds.), *Avaliação psicológica: Formas e contextos* (Vol. 6, pp. 504-526). Braga: Associação dos Psicólogos Portugueses (APPORT).
- 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. (2001a). 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. y Garrido, E. (2001b). Is there a kernel of truth in judgements of deceptiveness? *Anales de Psicología*, 17, 101-120.

- Masip, J. y Garrido, E. (2003). La detección del engaño mediante la técnica del control de la realidad. *Encuentros en Psicología Social*, 1, 113-121.
- Masip, J. y Garrido, E. (2007). *La evaluación del abuso sexual infantil*. Sevilla: MAD.
- Masip, J., Garrido, E. y Herrero, C. (2002a). 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. (2002b). 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. (2003a). *The history of lie detection: Early unscientific procedures*. Manuscrito no publicado.
- Masip, J., Garrido, E. y Herrero, C. (2003b). El Análisis de Contenido Basado en Criterios (CBCA). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 15, 75-92.
- Masip, J., Garrido, E. y Herrero, C. (2003c). When did you conclude she was lying? The impact of the moment the decision about the sender's veracity is made and the sender's facial appearance on police officers' credibility judgments. *Journal of Credibility Assessment and Witness Psychology*, 4, 1-36.
- Masip, J., Garrido, E. y Herrero, C. (2004a). Defining deception. *Anales de Psicología*, 20, 147-171.
- Masip, J., Garrido, E. y Herrero, C. (2004b). 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. (2006). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *International Journal of Psychology*, 41, 304-319.
- Masip, J., Garrido, E. y Herrero, C. (2007). *Heuristic versus systematic processing in detecting deception: Questioning the truth bias*. Manuscrito enviado para su publicación.
- Masip, J., Garrido, E., Herrero, C., Antón, C. y Alonso, H. (2006). Officers as lie detectors: Guilty before charged. En D. Chadee y J. Young (Eds.), *Current themes and perspectives in social psychology* (pp. 187-205). Mona, Jamaica: The University of the West Indies Press.

- Masip, J., Garrido, E. y Rojas, M. (2000, Abril). *Beliefs about indicators of deception and truthfulness in a specific situation*. Comunicación presentada en la 10th European Conference of Psychology and Law, Limassol, Chipre.
- Masip, J., Sporer, L. S., Garrido, E. y Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime and Law*, 11, 99-122.
- McCornack, S. A. y Levine, T. R. (1990). When lovers become leery: The relationship between suspicion and accuracy in detecting deception. *Communication Monographs*, 57, 219-230.
- Meltzer, B. (2003). Lying: Deception in human affairs. *International Journal of Sociology and Social Policy*, 23, 61-79.
- Menzel, E. W. (1986). Preface. En R. W. Mitchell y N. Thomson (Eds.), *Deception, perspectives on human and nonhuman deceit* (pp. 14-19). Albany, NY: State University of New York Press.
- 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. U. (1997). The effects of cognitive capacity and suspicion on truth bias. *Communication Research*, 24, 556-570.
- Mitchell, R. W. (1986). A framework for discussing deception. En R. W. Mitchell y N. Thomson (Eds.), *Deception, perspectives on human and nonhuman deceit* (pp. 3-40). Albany, NY: State University of New York Press.
- Mitchell, R. W. y Thomson, N. (1996). *Deception, perspectives on human and nonhuman deceit*. Albany, NY: State University of New York Press.
- Montaño, P. A. y Jimenez, J. J. (2001). Tutorial sobre redes neuronales artificiales: El perceptrón multicapa. *Revista Electrónica de Psicología*, 5.
- Montaño, P. A. y Jimenez, J. J. (2002). Tutorial sobre redes neuronales artificiales. Los mapas autoorganizados de Kohonen. *Revista Electrónica de Psicología*, 6.
- Montoya, V. (2003). El origen del lenguaje. *Revista Digital de la Universidad de Guadalajara*. Extraído el 17 Octubre del 2004 de <http://www.perspectivaciudadana.com/030426/culturadepaz05.html>
- Morgan, L. H. (1993) *Antropología. Lecturas*. Madrid: McGraw Hill.

- 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*. WA: The National Academies Press.
- Ninguna prueba apunta a ETA. (sin fechar). *ElPaís.com*. Extraído el 23 de Julio del 2008 de http://www.elpais.com/comunes/2007/juicio_11m/index.html
- Newman, M. L., Pennebaker, J. W., Berry, D. S. y Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.
- O'Hair, H. D. y Cody, M. J. (1994). Deception. En R. C. William y S. Brien (Eds.), *The dark side of interpersonal communication* (pp. 181-213). Hillsdale, NJ: Erlbaum.
- O'Sullivan, M. y Ekman, P. (2004). The wizards of deception detection. En P. A. Granhag, y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts*. Cambridge: Cambridge University Press.
- Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K. y Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, 69, 144-157.
- Pavlidis, I., Eberhardt, N. L. y Levine, J. A. (2002). Seeing through the face of deception. *Nature*, 415, 35-36.
- Pavlidis, I. y Levine, J. (2002a). Thermal facial screening for deception detection. *Proceedings of the Second Joint EMBS-BMES Conference* (Vol. 2, pp. 1143-1144). Houston, Texas.
- Pavlidis, I. y Levine J. (2002b). Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21, 56-64.
- Pennebaker, J. W., Francis, M. E. y Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahwah, NJ: Erlbaum.
- Peters, E. (1987) *La tortura*. Madrid: Alianza.
- Phan, K. L., Magalhaes, A., Ziemlewicz, T. J., Fitzgerald, D. A., Green, C. y Smith, W. (2005). Neural correlates of telling lies: A functional magnetic resonance imaging study at 4 tesla. *Academic Radiology*, 12, 164-172.
- Porter, S., Woodworth, M. y Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, 24, 643-658.

- Porter, S. y Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-458.
- Pitarque, A., Roy, J. F. y Ruiz, J. C. (1998). Redes neuronales vs. modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. *Psicológica*, 19, 387-400.
- Qin, T., Burgoon, J. K. y Nunamaker, J. F. (2004). An exploratory study on promising cues in deception detection and application of decision tree. *Proceedings of the 37th Hawaii International Conference on System Sciences*. Extraído el 1 del 12 del 2008 de <http://www2.computer.org/plugins/dl/pdf/proceedings/hicss/2004/2056/01/205610023b.pdf?template=1&loginState=1&userData=anonymousIP122814095693>
- Rae, G. (1976). Table of A'. *Perceptual and Motor Skills*, 42, 98.
- Raskin, D. C. (1994). *Métodos psicológicos en la investigación y pruebas criminales* (Gorbeña, S., Trad.). Bilbao: Desclée de Brouwer. (Trabajo original publicado en 1990).
- Raskin, D. C. y Esplin, P. W. (1991a). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, 13, 265-291.
- Raskin, D. C. y Esplin, P. W. (1991b). Assessment of children's statements of sexual abuse. En J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 153-164). Washington, DC: American Psychological Association.
- Real Academia Española. (2004). *Diccionario de la lengua española* (22ª Ed.). Madrid: Espasa Calpe.
- Richard, F. D., Bond, C. F. y Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331-363.
- Rico, J. M. (1988). *Inseguridad ciudadana y policía*. Madrid: Tecnos.
- Robinson, W. P. (1996). *Deceit, delusion, and detection*. Thousand Oaks, CA: Sage.
- Rodríguez, J. e Yoldi, J. (2004, 6 de Mayo). Acebes insistió en la tesis de ETA cuando la policía ya tenía cercados a los autores del 11-M. *El País*, pp. 18-19.
- Rosenfeld, J. P. (2002). Event-related potentials in the detection of deception, malignering, and false memories. En M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 265-286). San Diego, CA: Academic Press.

- Rosenfeld, J. P., Soskins, M., Bosh, G. y Ryan, A. (2004). Simple, effective countermeasures to P300-based test of detection of concealed information. *Psychophysiology*, 41, 205-219.
- Rothwell, J., Bandar, Z., O'Shea, J. y McLean, D. (2006). Silent talker: A new computer-based system for the analysis of facial cues to deception. *Applied Cognitive Psychology*, 20, 757-777.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651-665.
- 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.
- Seiter, S., Bruschke, J. y Chunsheng, B. (2002). The acceptability of deception as a function of perceivers' culture, deceiver's intention and deceiver – deceived relationship. *Western Journal of Communication*, 66, 158-180.
- Shearer, R. A. (1999). Statement analysis: SCAN or scam? *Skeptical Inquirer*, 23, 40-43.
- Smith, N. (2001). *Reading between lines: An evaluation of the scientific content analysis technique (SCAN)*. Londres: Home Office.
- Spence, S., Farrow, T., Herford, A., Wilkinson, L., Zheng, Y. y Woodruff, P. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, 12, 2849-2853.
- 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.
- Sporer, S. L. (2004). Reality monitoring and detection deception. En P. A. Granhag y L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 64-102). Cambridge: Cambridge University Press.
- Sporer, S. L., Bursch, S. E., Schreiber, N., Weiss, P. E., Höfer, E., Sievers, et al. (2000). Detecting deception with the Aberdeen Report Judgement Scales (ARJS): Inter-rater reliability. En A. Czerederecka, T. Jaskiewicz-Obydzinska y J. Wojcikiewicz (Eds.), *Forensic psychology and law. Traditional questions and new ideas* (pp. 197-204). Cracovia: Institute of Forensic Research Publishers.
- Sporer, S. L. y Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, 20, 421-446.

- Sporer, S. L. y Schwandt, B. (2007). Moderators of nonverbal indicators of deception. A meta-analytic synthesis. *Psychology, Public Policy, and Law*, 13, 1-34.
- Sporer, S. L. y Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology*, 20, 837-854.
- Stanislaw, H. y Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, 31, 137-149.
- 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 Boychuk, T. (1992). Children as witnesses in sexual abuse cases: Investigative interview and assessment techniques. En H. Dent y R. Flin (Eds.), *Children as witnesses*. (pp. 47-71). Chichester, Reino Unido: John Wiley and Sons.
- Steller, M. y Köhnken, G. (1994). Análisis de declaraciones basado en criterios. En D. C. Raskin (Ed.), *Métodos psicológicos en la investigación y pruebas criminales* (pp. 189-211). Bilbao: Desclée de Brouwer. (Trabajo original publicado en 1990).
- Stiff, J. B., Kim, H. J. y Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research*, 19, 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.
- Suriá, R. (2004, Octubre). Análisis de los foros virtuales: Utilización de un programa informatizado para conocer su contenido psicológico y lingüístico. *II Congreso online para el observatorio de la cibersociedad*. Extraído el 19 de Noviembre del 2007 de <http://www.monografias.com/trabajos901/analisis-foros-virtualespsicologico-linguistico/analisis-foros-virtuales-psicologico-linguistico.shtml#top>
- Swets, J. A., Dawes, R. B. y Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26.
- The Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, 37, 60-74.
- Tomás y Valiente, F. (1994). *La tortura en España* (2ª ed. revisada y aumentada), Barcelona: Ariel.
- Tomás y Valiente, F. (1979). *Manual de historia del derecho español*. Madrid: Tecnos.

- Toris, C. y DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology*, 47, 1063-1073.
- Tornqvist, S., Anderson, D. y DePaulo, M. (2001). Deceiving. En W. Peter Robinson y H. Giles (Eds.), *The new handbook of language and social psychology* (pp.271-284). Chichester, Reino Unido: John Wiley and Sons.
- Trovillo, P. V. (1939). A history of lie detection. *Journal of Criminal Law and Criminology*, 29, 848-881.
- Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I., Frank, M. G. y Ekman, P. (2007). Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, 71, 197-214.
- Tversky, A. y Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1134.
- Undeutsch, U. (1982). Statement reality analysis. En A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials*. (pp. 27-56). Estocolmo: Norsted and Soners.
- Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. *International Review of Applied Psychology*, 33, 51-67.
- Undeutsch, U. (1989). The development of statement reality analysis. En J. C. Yuille (Ed.), *Credibility assessment*. (pp. 101-119). Dordrecht, Holanda: Kluwer Academic Publishers.
- Veinte mentiras sobre la guerra de Irak. (2003, 20 de Octubre). *Infotk*. Extraído el 25 de Mayo del 2004 de http://infotk.blogs.com/infotk/2003/10/20_mentiras_sob.html
- Vrij, A. (1992). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of Social Psychology*, 133, 601-610.
- Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior*, 18, 117-136.
- Vrij, A. (1998a). Nonverbal communication and credibility. En A. Memon, A. Vrij y R. Bull. (Eds.), *Psychology and law. Truthfulness, accuracy and credibility* (pp. 32-58). Nueva York: McGraw-Hill.

- Vrij, A. (1998b). Physiological parameters and credibility: The polygraph. En A. Memon, A. Vrij y R. Bull. (Eds.), *Psychology and law. Truthfulness, accuracy and credibility* (pp. 77-101). Nueva York: McGraw-Hill.
- Vrij, A. (2000). *Detecting lies and deceit. The psychology of lying and the implications for professional practice*. Chichester, Reino Unido: John Wiley and Sons.
- Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology*, 9, 159-181.
- Vrij, A. (2005). Criteria-based content analysis. A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3-41.
- Vrij, A. y Akehurst, L. (1998). Verbal communication and credibility: Statement validity assessment. En A. Memon, A. Vrij y R. Bull. (Eds.), *Psychology and law. truthfulness, accuracy and credibility*. Nueva York: McGraw Hill.
- Vrij, A., Akehurst, L., Brown, L. y Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology*, 20, 1225-1237.
- Vrij, A., Akehurst, L., Soukara, S. y Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8-41.
- Vrij, A., Edward, K., Roberts, K. y Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263.
- Vrij, A., Fisher, R., Mann, S. y Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10, 141-142.
- Vrij, A. y Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence*, 5, 144-148.
- Vrij, A. y Mann, S. (2001a). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology*, 15, 187-203.
- Vrij, A., y Mann, S. (2001b). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime and Law*, 7, 119-132.
- Vrij, A., Mann, S. y Fisher, R. (2006). An empirical test of the Behaviour Analysis Interview. *Law and Human Behavior*, 30, 329-345.
- Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, R. y Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, 32, 253-265.

- Vrij, A., Mann, S., Kristen, S. y Fisher, R. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31, 499-518.
- Vrij, A., Mann, S., Robbins, E. y Robinson, M. (2006). Police officers ability to detect deception in high stakes situations and in repeated lie detection test. *Applied Cognitive Psychology*, 20, 741-755.
- Vrij, A. y Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, 20, 65-80.
- Vrij, A. y Winkel, F. W. (1993). Objective and subjective indicators of deception. *Issues in Criminological and Legal Psychology*, 20, 51-57.
- Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 36, 1546-1557.
- Wrightman, L. (1964). Measurement of philosophies of human nature. *Psychological Reports*, 14, 743-751.
- 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.
- Zimbardo, P. G., Haney, C., Banks, W. C. y Jaffe, D. (1973, 8 de abril). The mind is a formidable jailer: A pirandellian prison. *The New York Times Magazine*. Extraído el 1 del 12 de <http://www.prisonexp.org/pdf/pirandellian.pdf>
- Zhou, L., Burgoon, J. K., Nunamaker, J. F. y Twitchell, D. P. (2004). Automating linguistic-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation*, 13, 81-106.
- Zhou, L. y Zhang, D. (2006). A comparison of deception behavior in dyad and triadic group decision making in synchronous computer-mediated communication. *Small Group Research*, 37, 140-164.
- Zuckerman, M., DePaulo, B. M. y Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, 14, 1-59.

ANEXOS

ANEXO A

LISTA DE VALIDEZ TAL Y COMO SE PRESENTA EN MASIP Y GARRIDO (2007)

I. FACTORES RELACIONADOS CON LA DECLARACIÓN

A) Características psicológicas

1. Limitaciones cognitivo-emocionales
2. Lenguaje y conocimiento
3. Emociones durante la entrevista
4. Sugestionabilidad

B) Características de la entrevista

5. Procedimientos de entrevista
6. Influencia sobre los contenidos de las declaraciones

C) Motivación

7. Circunstancias de la alegación original
8. Motivación para declarar
9. Influencia por parte de otros

II. CUESTIONES INVESTIGATIVAS

10. Falta de realismo
11. Declaraciones inconsistentes
12. Evidencia contradictoria
13. Características del delito

ANEXO B.

LAS OCHO ESCALAS DEL JMCQ. DATOS TOMADOS DE SPORER (2004)

1. Claridad y viveza

Claridad con que se recuerda el evento.

Cantidad de detalles visuales.

Viveza con la que se describe el evento.

Precisión de los detalles.

Orden de los eventos.

Calidad del recuerdo.

2. Información sensorial

Descripción de **colores** de los elementos del evento.

Descripción de **sonidos** escuchados durante el evento.

Descripción de **olores** recordados durante el evento.

Descripción del recuerdo del **tacto** de objetos tocados.

Descripción de **sabores** recordados durante el evento.

3. Información espacial

Localización del evento recordado.

Familiaridad del declarante con el **ambiente / situación** del evento.

Descripción de la **disposición de los objetos** en el espacio.

Descripción de la **disposición de las personas** en el espacio.

4. Información temporal

Claridad de la descripción del **momento** en que ocurrió el evento.

Claridad del recuerdo del **año** en que ocurrió el evento recordado.

Claridad del recuerdo de la **estación** en que ocurrió el evento.

Claridad del recuerdo del **día** en que ocurrió el evento recordado.

Claridad del recuerdo de la **hora** en que ocurrió el evento recordado.

5. Emociones y sentimientos

Papel / rol desempeñado en el evento recordado.

Claridad del **recuerdo de sentimientos** durante el evento.

Intensidad de los sentimientos vividos durante el evento.

Intensidad de los sentimientos ahora, al recordar el evento.

Aspectos de la personalidad que revela la historia recordada.

6. Reconstruibilidad de la historia

Complejidad del hilo conductor de la historia recordada.

Consideración de las consecuencias del evento mientras ocurría.

Consecuencias reales que tuvo el evento tras haber ocurrido.

Dudas sobre la precisión de los recuerdos.

ANEXO B (CONTINUACIÓN)

7. Realismo

Realismo de la historia.

Posibilidad de suceder(me) este evento en una situación similar.

Aspectos extraordinarios, sorprendentes, impredecibles del evento.

Detalles increíbles de la historia.

Credibilidad general, global de la historia

8. Operaciones cognitivas

* Se incluye el mismo tipo de información que para la escala Reconstruibilidad del a historia, pero haciendo referencia a pensamientos, razonamientos o inferencias sobre el evento.

ANEXO C.

ESCALAS DE LAS ARJS (SPORER, 2004).

Escalas / Criterios	Ítems
Características globales	
Escala 1. Realismo y coherencia	1-4
Escala 2. Claridad y viveza	5-8
Cantidad y precisión de los detalles	
Escala 3. Detalles centrales y periféricos y precisión de los detalles	9-12
Escala 4. Información espacial	13-16
Escala 5. Información temporal	17-21
Escala 6. Impresiones sensoriales	22-25
Procesos internos	
Escala 7. Emociones y sentimientos	26-29
Escala 8. Pensamientos y procesos cognitivos	30-32
Escala 9. Procesos de memoria y ensayo	34-38
A. Ensayo interno y externo (ítems 34 y 35)	
B. Memorias de apoyo (ítems 36 y 38)	
Aspectos sociales	
Escala 10. Interacciones verbales y no verbales	39-43
Escala 11. Detalles extraordinarios	44-45
Escala 12. Falta de de deseabilidad social	33, 46-48
Memoria autobiográfica	
Escala 13. Significación personal e implicaciones	49-52

ANEXO D

CRITERIOS MÁS COMÚNMENTE UTILIZADOS EN LA TÉCNICA SCAN Y SUS DEFINICIONES. INFORMACIÓN DE TOMADA DE MASIP ET AL. (2002b) Y ADAMS (1996).

1. Empleo de pronombres

El pronombre yo: se espera que las declaraciones verdaderas se presenten en primera persona (norma) una desviación de esta norma debe alertar de la posibilidad de engaño.

El pronombre nosotros: se espera que se emplee cuando la historia se refiere a dos personas que deberían estar muy unidas (esposos, padre e hijo). Si en los casos en los que es esperable que aparezca no lo hace, puede ser signo de engaño.

Pronombres posesivos: la ausencia de un pronombre posesivo donde debería aparecer indica posibilidad de engaño, ya que el individuo tratará de no mostrar su relación o asociación con el objeto poseído.

2. Presentación social

La referencia a cada uno de los personajes que aparece en la declaración debe ser clara y sin ambigüedades. Un cambio en la forma de referirse a un personaje (ej., pasar de utilizar su nombre propio a utilizar un pronombre) indica posibilidad de engaño.

3. Correcciones espontáneas

Cualquier corrección que aparezca en la declaración (ej., palabras tachadas) es un signo de engaño. Cuando se pide al individuo que escriba su declaración, se le indica que no debe hacer correcciones.

4. Distribución temporal de los hechos

Se espera que la declaración tenga tres partes: antes del hecho principal, el hecho principal y después del hecho principal. Que las tres partes no sean igual de largas se interpreta como signo de engaño.

5. Tiempo subjetivo

El tiempo empleado en describir un hecho ha de ser proporcional a la duración real del hecho (ej., si un declarante emplea tres líneas en describir un hecho que en tiempo real ha durado diez minutos, ha de emplear seis líneas en describir un hecho que haya durado veinte minutos) de lo contrario, se sospecha engaño.

6. Localización de las emociones

Se afirma que en las declaraciones verdaderas la mayor tasa de emocionalidad aparece justo después de la parte culminante de la declaración. De no ser así, se debe sospechar la falsedad de la historia.

ANEXO D (CONTINUACIÓN)

7. Conectores y espacios vacíos en la narración

Se entiende por conector un elemento que une dos partes de la declaración (“luego”, “seguidamente”, “después”). Se indica que la presencia de conectores implica que hay periodos de la declaración en los que falta información. A más conectores, más posibilidad de que las declaraciones sean falsas.

8. Falta de memoria

Afirmaciones, por parte del declarante, de falta de memoria, indican engaño.

9. Enlaces ausentes o enlaces innecesarios

La presencia de enlaces ausentes (información sin la que la historia no puede continuar) sugiere engaño. Los enlaces innecesarios (información que se puede suprimir de la declaración sin que afecte a su línea argumental) sugieren que la historia es cierta.

10. Empleo del tiempo verbal pasado

Dado que la declaración versa sobre un hecho que ya ha sucedido, debe aparecer en pasado. Un cambio al tiempo verbal presente sugiere engaño.

11. Negación de alegaciones

El individuo inocente, que dice la verdad, negará directamente su culpabilidad: “yo no lo hice”. El culpable no. Las negaciones directas indican veracidad.

12. Lógica de la narración

Se espera que la mayor parte de las declaraciones falsas sean lógicas. Las verdaderas no tienen porqué serlo.

13. Inconsistencia entre alegaciones

Cualquier inconsistencia entre alegaciones es signo de engaño.

14. Cambios en el lenguaje

Los cambios de palabras que han aparecido de forma consistente a lo largo de la declaración sugieren engaño.

15. Hacer preguntas al entrevistador

Cuestionar las preguntas que pueda hacer el entrevistador sugiere engaño.

16. Cronología y concisión

Las declaraciones verdaderas tienen un orden cronológico que se corresponde con el orden en el que realmente sucedieron los hechos.

ANEXO D (CONTINUACIÓN)

17. Inseguridades verbales

Expresiones que denotan inseguridad (“creo”, “pienso”), indican engaño.

ANEXO E

ESCALA DE GCS ORIGINAL ELABORADA POR LEVINE Y MCCORNACK (1991),
NÚMERO DE ÍTEMS FINALMENTE RETENIDOS

1. Everyone lies, the person who says that they don't is the biggest liar of all.
2. I often feel as if people aren't being completely truthful with me.
3. Most people only tell you what they think you want to hear.
4. When i am in a conversation with someone, I frecuently wonder whether they are really telling me the truth.
5. People rarely tell you what they're really thinking.
6. The best policy is to trust people until proven wrong.
7. Dishonesty is a part of human nature.
8. When I first meet someone, I asume that they are probably lying to me about some things.
9. Most people are basically honest.
10. Anyone who completely trusts someone else is asking for trouble.
11. When I ask a stranger for directions, I asume that they are probably lying to me about some things.
12. When I am talking to others, I tend to believe what they say.
13. People seldom lie to me.
14. Most people follow the saying "honesty is the best policy".

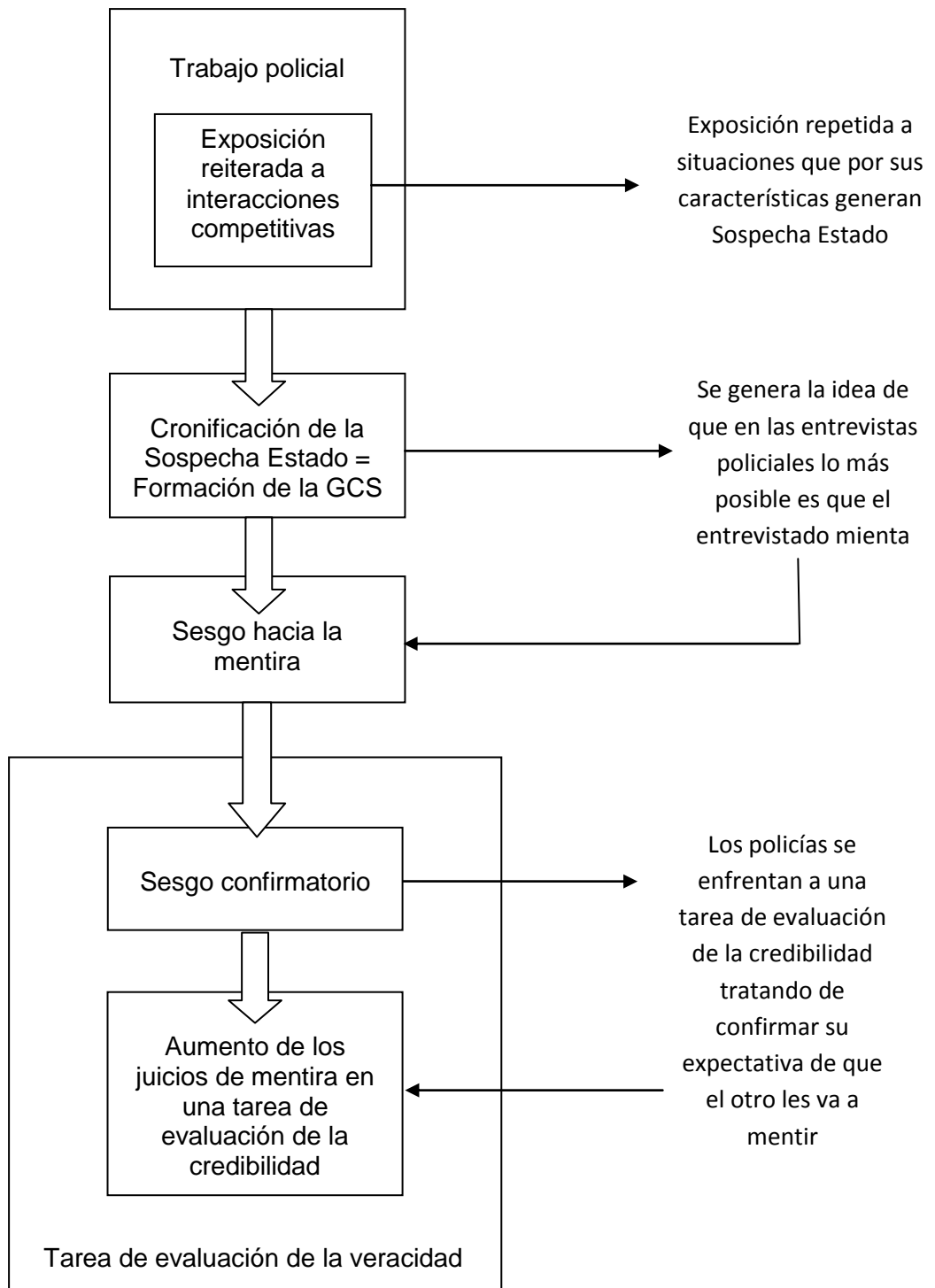
Los ítems 2, 3, 4, 5, 8, 9, 10, 11, 12, 13 y 14 se retuvieron en el estudio de Levine y McCornack (1991) y formaron la escala definitiva.

Cada ítem se contestó en una escala tipo *Likert* de siete puntos de 1 (*fuertemente de acuerdo*) a 7 (*fuertemente en desacuerdo*).

La puntuación final se obtuvo sumando la puntuación de cada uno de los ítems, pero teniendo en cuenta que las respuestas a los ítems 6, 9, 12, 13 y 14 tenían que codificarse de forma inversa.

ANEXO F

ESQUEMA DE LA FORMACIÓN DEL SESGO HACIA LA MENTIRA DE LOS POLICÍAS BASADO EN LA GCS



ANEXO G

LOS 14 ÍTEMS DE LA ESCALA DE GCS ELABORADA POR LEVINE Y MCCORNACK (1991) TRADUCIDOS AL ESPAÑOL.

1. Todo el mundo miente, y quien diga que no lo hace es el mayor mentiroso de todos.
2. A menudo me parece que la gente no es del todo veraz conmigo.
3. La mayoría de la gente sólo te dice lo que ellos creen que quieres oír.
4. Cuando estoy conversando con alguien, a menudo me pregunto si realmente me está diciendo la verdad.
5. Es sólo en raras ocasiones que la gente te cuenta lo que está pensando realmente.
6. La mejor estrategia consiste en confiar en la gente hasta que se demuestre lo contrario.
7. La falta de veracidad es parte de la naturaleza humana.
8. Cuando conozco a alguien, asumo que probablemente me está diciendo algunas mentiras.
9. La mayoría de las personas son básicamente veraces.
10. Quien confíe ciegamente en otra persona se está buscando problemas.
11. Cuando estoy en una ciudad extraña y pregunto a un desconocido por el sitio que estoy buscando, a menudo me pregunto si no me estará engañando.
12. Cuando hablo con otras personas tiendo a creer lo que dicen.
13. La gente me miente sólo en raras ocasiones.
14. La mayoría de las personas sigue la máxima que dice "la mejor estrategia es decir la verdad".

Los ítems 2, 3, 4, 5, 7, 8, 9, 10, 12, 13 y 14 se retuvieron en nuestro estudio y formaron la escala definitiva.

Cada ítem se contestó en una escala tipo *Likert* de siete puntos de 1 (*fuertemente de acuerdo*) a 7 (*fuertemente en desacuerdo*).

La puntuación final se obtuvo sumando la puntuación de cada uno de los ítems, pero teniendo en cuenta que las respuestas a los ítems 9, 12, 13 y 14 tenían que codificarse de forma inversa.

ANEXO H

ESCALAS, CATEGORÍAS Y FORMAS DE ACCESO DENTRO DEL CUERPO NACIONAL DE POLICÍA DE ESPAÑA.

ESCALA BÁSICA

Comprende las categorías de Policía a las que se accede por oposición libre reuniendo los siguientes requisitos:

- a) Ser español.
- b) Tener 18 años de edad y no haber cumplido los 30.
- c) Estatura mínima de 1.65 metros los hombres y 1.60 las mujeres.
- d) Compromiso de portar armas y, en su caso, llegar a utilizarlas.
- e) No haber sido condenado por delito doloso ni separado del servicio del Estado, de la administración Autonómica, Local o Institucional, ni hallarse inhabilitado para el ejercicio de funciones públicas.
- f) Permiso de conducir de la clase que se determine en la respectiva convocatoria.
- g) Estar en posesión o en condiciones de obtener el título de Graduado en Secundaria o formación equivalente o superior.

También se incluye en la Escala Básica la categoría de Oficial de Policía, a la que se accede por concurso oposición interna.

ESCALA DE SUBINSPECCIÓN

Comprende la categoría de subinspector, a la que se accede por antigüedad selectiva en el Cuerpo o a través de concurso oposición interno.

ESCALA EJECUTIVA

Comprende la categoría de Inspector, a la que se accede por antigüedad selectiva en el Cuerpo, concurso oposición interno u oposición libre reuniendo los siguientes requisitos:

- a) Ser español.
- b) Tener 18 años de edad y no haber cumplido los 30.
- c) Estatura mínima de 1.65 metros los hombres y 1.60 las mujeres.
- d) Estar en posesión del título de Ingeniero técnico, Arquitecto técnico, Diplomado universitario o equivalente.
- e) Compromiso de portar armas y, en su caso, llegar a utilizarlas.

ANEXO H (CONTINUACIÓN)

- f) No haber sido condenado por delito doloso ni separado del servicio del Estado, de la administración Autonómica, Local o Institucional, ni hallarse inhabilitado para el ejercicio de funciones públicas.
- g) Permiso de conducir de la clase que se determine en la respectiva convocatoria.

Esta Escala comprende también la Categoría de Inspector jefe, a la que se accede por antigüedad selectiva o concurso oposición interna.

ESCALA SUPERIOR

Comprende las categorías de Comisario y Comisario principal, a las que se accede por antigüedad selectiva o concurso oposición interna.

ANEXO I

INSTRUCCIONES LEÍDAS A LOS PARTICIPANTES PARA LA CUMPLIMENTACIÓN DEL CUESTIONARIO DE GCS.

“A continuación, se les presentará un ejercicio en el que se les va a plantear una serie de cuestiones a las que deben mostrar su parecer”.

“En concreto, se les va a entregar un cuestionario en el que se les pedirá que muestren su grado de acuerdo o desacuerdo con cada una de las afirmaciones propuestas, reflejándolo en una escala que aparece en la parte inferior de cada ítem, en la que el 1 implicaría que están totalmente en desacuerdo con la afirmación propuesta, y el 7 implicaría que están totalmente de acuerdo. Les recordamos que la tarea ha de realizarse individualmente”.

DATOS DEL CUESTIONARIO

“En la cabecera del cuestionario encontrarán ustedes una serie de datos que es importante que se cumplimenten de forma correcta, así que les pido que, antes de comenzar a responder a las preguntas del cuestionario en sí, rellenen estos datos. Deben indicar el sexo, el modo de acceso a la Escala Ejecutiva, la fecha de nacimiento (es importante que se refleje el día mes y año) y los años de experiencia. Además, en la parte superior derecha, deben indicar el grupo o aula a la que pertenecen, (A, B, C...), y los que tengan años de experiencia y procedan de la básica, deben también indicarlo”.

“¿Alguna pregunta sobre la cumplimentación de los datos iniciales del cuestionario?”

“Si alguien tiene alguna duda relacionada con el objetivo de la investigación, se abrirá un turno de preguntas tras la cumplimentación del cuestionario”.

ANEXO J

INSTRUCCIONES CORRESPONDIENTES A LA TAREA DE EVALUACIÓN DE LA CREDIBILIDAD DADAS A LOS POLICÍAS (ESTUDIO 3)

INSTRUCCIONES PRESENTADAS POR ESCRITO Y ADEMÁS LEÍDAS POR EL INVESTIGADOR

A continuación van a ver ustedes a 12 personas haciendo una declaración.

Cada una de estas personas ha visto un vídeo, y luego ha tenido 10 minutos para preparar su declaración sobre lo que sucedía en el vídeo.

A cada una se le ha hecho tres preguntas sobre lo que sucedía en el vídeo. Ustedes van a ver las tres respuestas de cada testigo a tales preguntas, aunque las preguntas se han suprimido de la cinta.

La declaración de cada una de estas personas puede ser verdadera o puede ser falsa. Es decir, puede que haya dicho la verdad al contestar a las tres preguntas, o puede que haya mentido. Deben ustedes tener en cuenta que si dice la verdad la dice en las tres respuestas, y si miente lo hace también en sus tres respuestas.

La tarea que solicitamos de ustedes consiste en:

Observar los vídeos de las declaraciones y determinar si cada uno de los testigos miente o dice la verdad.

El juicio sobre cada testigo deben hacerlo ustedes después de haber visto al testigo respondiendo a las tres preguntas. En ese momento el experimentador parará el vídeo para que tengan tiempo de responder.

Además de indicar si creen que el testigo miente o dice la verdad, deberán señalar también su grado de seguridad en ese juicio en una escala que va de 0 a 7.

Estas preguntas se formulan en el cuestionario que se les ha entregado.

¿Han comprendido ustedes la tarea?

INSTRUCCIONES PRESENTADAS SOLO ORALMENTE POR EL INVESTIGADOR

Hay algunas cosas que deben tener Vds. en cuenta al efectuar los juicios:

- ✓ Había varios vídeos: No todos los testigos han visto el mismo vídeo.*
- ✓ En la mayoría de ocasiones –aunque no siempre– los vídeos eran bastante parecidos entre ellos.*
- ✓ En la mayoría de ocasiones –aunque no siempre– los sujetos mentirosos no alteraban radicalmente lo que habían visto en el vídeo, sino que introducían pequeños cambios.*

ANEXO J (CONTINUACIÓN)

Esto hace que el contenido de las declaraciones verdaderas y falsas a menudo sea muy parecido, y dificulta la tarea. Es importante que a la hora de juzgar las declaraciones no comparen unas con otras, sino que juzguen cada una de ellas en sí misma.

En cualquier caso, en todas las versiones de los vídeos aparecían 3 personajes:

- ✓ *Uno de sexo femenino*
- ✓ *Dos de sexo masculino:*
 1. *Un hombre con bigote*
 2. *Un hombre con traje*

Las preguntas que se hicieron a los testigos fueron:

- ✓ *describe lo que hizo la mujer*
- ✓ *describe lo que hizo el hombre del bigote*
- ✓ *describe lo que hizo el hombre del traje*

aunque no siempre en el mismo orden.

¿Han comprendido Vds. la tarea?

ANEXO K

LOS 13 ÍTEMS QUE COMPONEN LA ESCALA DE SOSPECHA EN INTERROGATORIOS, ELABORADA A PARTIR DE LA ESCALA ORIGINAL DE GCS DE LEVINE Y MCCORNACK (1991).

1. Todo sospechoso en un interrogatorio miente, y el que diga que no lo hace es el mayor mentiroso de todos.
2. A menudo me parece que en mi trabajo, cuando pregunto a un sujeto información que de alguna forma podría estar relacionada con un delito, éste no es del todo veraz conmigo.
3. Cuando estoy interrogando a alguien, a menudo me pregunto si realmente me está diciendo la verdad.
4. Sólo en raras ocasiones, la gente cuenta lo que está pensando realmente cuando tiene que responder a las preguntas que les formula un policía, sobre su posible implicación en determinados hechos delictivos.
5. La mejor estrategia como policía, consiste en confiar en los sujetos a los que se interroga hasta que se demuestre lo contrario.
6. La falta de veracidad es parte de la naturaleza humana cuando la gente se ve sometida a un interrogatorio policial.
7. Cuando conozco a alguien, porque ya le he interrogado en otras ocasiones, asumo que probablemente me está diciendo algunas mentiras.
8. La mayoría de las personas son básicamente veraces cuando un policía les interroga sobre determinados hechos de los que son sospechosos.
9. El policía que confíe ciegamente en la declaración de un sospechoso, se está buscando problemas.
10. Cuando un policía está de servicio y pregunta a un desconocido por un sitio que está buscando, a menudo debería preguntarse si no me estará engañando.
11. Cuando interrogo a un sospechoso, tiendo a creer lo que me dice.
12. La gente que tengo que interrogar, me miente sólo en raras ocasiones.
13. La mayoría de los sujetos que saben que se les va a interrogar, siguen la máxima que dice "la mejor estrategia es decir la verdad".

Los ítems 1, 2, 3, 4, 5, 6, 7, 8, 11, 12 y 13 se retuvieron y formaron la escala definitiva.

Cada ítem se contestó en una escala tipo *Likert* de siete puntos de 1 (*fuertemente de acuerdo*) a 7 (*fuertemente en desacuerdo*).

La puntuación final se obtuvo sumando la puntuación de cada uno de los ítems, pero teniendo en cuenta que las respuestas a los ítems 5, 8, 12 y 13 tenían que codificarse de forma inversa.

ANEXO L

EJEMPLOS DE LAS PREGUNTAS DE LA BAI, Y LAS REACCIONES QUE SE ESPERAN EN LOS INOCENTES Y LOS CULPABLES (TABLA-RESUMEN ELABORADA A PARTIR DEL TEXTO DE INBAU ET AL., 2001).

Pregunta	Ejemplo	Reacciones del Inocente :	Reacciones del Culpable :
1. Objetivo	Jim, ¿Cuál crees que es el objetivo de esta entrevista conmigo hoy aquí?	<ul style="list-style-type: none"> • Respuesta directa. Por ej., “Estoy seguro de que usted quiere saber qué es lo que yo sé sobre el incendio en el trabajo • Lenguaje realista. Por ej., usar la palabra “incendio” 	<ul style="list-style-type: none"> • Respuesta ingenua o evasiva • Comentario vago. Por ej., “Supongo que usted quiere hablar conmigo de lo que sucedió en el almacén”
2. Historia / Usted	Jim, como ya sabes, hace dos días que hubo un incendio en el almacén de la calle 6. Si tú provocaste ese fuego, nuestra investigación nos va a permitir descubrirlo. Si no tuviste nada que ver, también podemos demostrarlo. Si tuviste algo que ver deberías decírmelo ahora. (La cuestión también podría formularse haciendo la pregunta directa “¿Provocaste tú ese fuego?”).	<ul style="list-style-type: none"> • Negación empática e inmediata. Por ej., “de ningún modo. Yo no tengo absolutamente nada que ver con el inicio del fuego” • Conductas no verbales como: <ul style="list-style-type: none"> - Inclinarsse hacia delante en la silla. - Establecer contacto visual directo. - Usar ilustradores 	<ul style="list-style-type: none"> • Respuesta ingenua o evasiva. Por ej., - “Por Dios Santo, yo no lo hice; lo juro”. - “¿Si lo provoqué yo? No, no lo hice. - “Es ahí donde trabajo, ¿por qué haría algo así?” • Conductas no verbales como cruzar las piernas, cambiar de postura en la silla o acicalamiento
3. Conocimiento	Jim, ¿sabes quien provocó ese fuego?	<ul style="list-style-type: none"> • Insinúa una sospecha. Por ej., “Bueno, no lo sé con seguridad, se han dicho algunos nombres en el trabajo, pero no lo sé con certeza” • Negación acompañada de disculpa sincera: “Ojalá lo supiera, pero no tengo ni la más mínima idea” • Suena sincero al dar la respuesta • Señala que ha estado pensando sobre quien puede ser el culpable del delito 	<ul style="list-style-type: none"> • Se distancia geográficamente y emocionalmente del delito • Niega, sin pensarlo mucho, cualquier conocimiento sobre quien pueda ser el culpable. Por ej., “No, no lo sé”, o un simple y rápido “No” • Respuesta evasiva. Por ej., “Ni siquiera sé si el fuego fue provocado”
4. Sospecha	Jim, ¿quién sospechas que puede haber provocado ese fuego? Deja que te diga que una sospecha puede ser sólo un vago presentimiento de tu parte, y que puede ser completamente errónea. Pero he visto que, en situaciones como esta, la mayoría de las personas sospechan de otros. Nadie sabrá que me has dado su nombre. ¿Quién crees que puede haber hecho esto?	<ul style="list-style-type: none"> • Nombra a una o más personas de quienes sospecha que posiblemente cometieron el delito • Afirma que sospecha de otro u otros. Es capaz de dar razones creíbles 	<ul style="list-style-type: none"> • Dice no tener ninguna sospecha sobre quien puede ser el culpable • Si la oportunidad de cometer el delito se limita a dos personas, afirma que sospecha del otro • Afirma que sospecha del otro, pero no logra dar ninguna razón creíble

ANEXO L (CONTINUACIÓN)

5. Garantía	Jim, de las personas que trabajan en el almacén, ¿hay alguien que estés seguro que no provocó el incendio, alguien por quien pondrías la mano en el fuego?	<ul style="list-style-type: none">• Nombra sin problema a personas concretas que cree que son irreprochables o que garantiza que no estuvieron implicadas en el delito	<ul style="list-style-type: none">• Respuesta evasiva. Por ej., “en realidad no... No conozco a la gente tan bien”.• Respuesta poco comprometida. Por ej., “pondría la mano en el fuego por cualquiera”
6. Credibilidad	Jim, ¿crees que realmente hay alguien que provocó deliberadamente ese incendio?	<ul style="list-style-type: none">• Reconoce que el delito ha tenido lugar. Por ej., “sí, lo creo. El fuego empezó justo en medio del pasillo. No hay cables eléctricos en esa zona, ni ninguna otra cosa que pudiera haber causado el fuego de forma accidental”	<ul style="list-style-type: none">• Sugiere posibilidades poco realistas (como un origen eléctrico para el incendio, o un uso poco cuidadoso de los materiales para fumadores)
7. Oportunidad	Jim, ¿quién habría tenido la mejor oportunidad de provocar ese incendio si hubiera querido hacerlo? No digo que esa persona lo hiciera, pero ¿quién podría haber tenido la posibilidad de provocar el fuego?	<ul style="list-style-type: none">• Admite que él mismo puede haber cometido el delito. Por ej.,<ul style="list-style-type: none">- “Cualquiera de nosotros que tenga llaves del almacén”.- “Probablemente un trabajador del tercer turno, como yo mismo y otras cinco personas”	<ul style="list-style-type: none">• Nombra a ciertas personas de manera poco realista. Por ej., “caramba, podría haber sido cualquiera de los que trabajan ahí, o cualquier antiguo trabajador. Incluso puede haber sido alguien de fuera de la empresa”.• Dice que nadie ha tenido una oportunidad de cometer el delito. Por ej., “por la noche cierran el almacén. No creo que nadie tuviera la oportunidad de provocar ese incendio”
8. Actitud	Jim, ¿Cómo te sientes al ser entrevistado sobre ese incendio?	<ul style="list-style-type: none">• Actitud positiva. Por ej.,<ul style="list-style-type: none">- “No me importa en absoluto... Haré lo necesario para llegar al fondo de esto”.- “Estoy contento de cooperar como sea, quiero que dejen de sospechar de mí”	<ul style="list-style-type: none">• Actitud negativa. Por ej.,<ul style="list-style-type: none">- “Me siento como un criminal”.- “Estoy nervioso y asustado”, o- “No comprendo por qué me hace estas preguntas, un montón de gente pudo haberlo hecho”
9. Pensamiento	Jim, ¿alguna vez has pensado sobre hacer algo para desquitarte de la empresa?	<ul style="list-style-type: none">• Niega con firmeza, de forma inequívoca, haber tenido estos pensamientos o ideas. Por ej., “en absoluto, no”	<ul style="list-style-type: none">• Dispuesto a admitir haber pensado sobre cometer el delito. Por ej., “bueno, claro. Seguro que la mayoría de los trabajadores tienen pensamientos como este de vez en cuando”.• Matiza su respuesta. Por ej., “realmente no” o “ en serio no”

ANEXO L (CONTINUACIÓN)

10. Motivo	Jim, ¿por qué crees que alguien provocó ese incendio?	<ul style="list-style-type: none">• Da un motivo razonable para el delito (en el caso del incendio, una respuesta razonable sería la ira, una venganza o un fraude a la compañía de seguros)• Parece cómodo al discutir posibles razones	<ul style="list-style-type: none">• Poco dispuesto a especular sobre los posibles motivos, respondiendo cosas como: “¿Cómo quiere usted que lo sepa” Puede haber cualquier razón”• “No tengo ni idea”• “No he pensado mucho en esto”• Respuesta muy detallada o específica. Por ej., “quizás esa persona fuera tratada de manera injusta por la empresa y estuviera borracha, así que hizo esto para desquitarse”• Durante su respuesta, cambia de postura en la silla o realiza otras conductas reductoras de ansiedad
11. Castigo	Jim, ¿qué crees que se debería hacer con la persona que provocó este incendio?	<ul style="list-style-type: none">• Sugiere tratamiento razonablemente duro. Por ej.,<ul style="list-style-type: none">- “Se le debería meter en la cárcel”.- “Espero que le juzguen y que lo manden a la cárcel”- “Debería ser despedido”	<ul style="list-style-type: none">• Sugiere tratamiento más bien indulgente. Por ej.,<ul style="list-style-type: none">- “Bueno, supongo que depende de las circunstancias”- “Sin duda necesitaría apoyo psicológico”• Se abstiene de proponer un castigo y responde, por ej.,<ul style="list-style-type: none">- “Esto no es cosa mía”- “Esto debería decidirlo un juez”
12. Segunda oportunidad	Jim, ¿hay alguna circunstancia bajo la cual crees que debería darse una segunda oportunidad a la persona que provocó ese incendio?	<ul style="list-style-type: none">• Opina que no hay que ofrecer una segunda oportunidad al culpable. Por ej.,<ul style="list-style-type: none">- “De ningún modo. Después de lo que ha pasado, espero que le caiga una buena”- “¡Joder, no! Quien haya hecho esto ha hecho que el almacén cerrara; estoy perdiendo dinero por culpa de ese incendio”	<ul style="list-style-type: none">• Está de acuerdo sobre dar una segunda oportunidad al culpable• Ofrece respuesta evasiva, Por ej., “es difícil saberlo...”• Alude a condiciones o circunstancias. Por ej., “bueno, creo que es importante averiguar cuáles fueron las circunstancias que le llevaron a eso”

ANEXO L (CONTINUACIÓN)

13. Objeción	Jim, dime porqué tu no harías algo así.	<ul style="list-style-type: none">• Menciona rasgo personal en respuesta en primera persona. Por ej.,<ul style="list-style-type: none">- “¡Porque yo no soy un pirómano!”.- “¡Nunca más podría dormir tranquilo si hubiera hecho algo así”.• Alude a sus responsabilidades presentes o logros pasados. Por ej., “nunca arriesgaría todo aquello por lo que he estado trabajando haciendo algo así”	<ul style="list-style-type: none">• Ofrece respuesta en tercera persona. Por ej.,<ul style="list-style-type: none">- “Esto es ilegal”- “Está mal”• Hace referencia a consecuencias futuras, como no querer ir a la cárcel o evitar hacer algo que pudiera constituir una violación de la libertad condicional• Alude a factores externos. Por ej., “tienen cámaras de seguridad por todos lados; me pillarían”
14. Resultados	Jim, en cuanto acabemos la investigación, ¿cuáles crees que van a ser los resultados respecto a tu participación en el origen de ese incendio?	<ul style="list-style-type: none">• Confía en que se demuestre su inocencia. Por ej.,<ul style="list-style-type: none">- “¡Van a mostrar que yo no tuve nada que ver con eso!”- “¡Yo sé que no provoqué el incendio, así que no estoy preocupado”- “Se va a demostrar que estoy diciendo la verdad sobre todo lo que te he dicho”	<ul style="list-style-type: none">• Responde con una sola palabra. Por ej., “limpio”, “inocente” o “bien”• Sentimiento de incertidumbre. Por ej.,<ul style="list-style-type: none">- “Espero que acabe bien”- “No tengo ni idea, supongo que ya lo veremos”• Respuesta evasiva. Por ej., “bueno, de hecho no tengo ningún control sobre tu investigación, así que no lo sé”• Anticipa resultados negativos, pero culpa a otra persona o cosa. Por ej.,<ul style="list-style-type: none">- “Siempre me culpan de cosas que no he hecho; seguro que esto no será una excepción”- “La verdad es que soy una persona nerviosa y la gente siempre piensa que estoy mintiendo aunque esté diciendo la verdad”
15. Contarlo a los Seres Queridos	¿A quién le has dicho algo sobre tu entrevista conmigo hoy?	<ul style="list-style-type: none">• Admite haber contado todos los detalles del tema de la investigación a algún ser querido	<ul style="list-style-type: none">• Dice que no ha dicho nada de la entrevista a ningún ser querido• Ha quitado importancia a la entrevista al hablar con algún ser querido• Si se le pregunta cómo reaccionó el ser querido, responde algo como: “ bueno, ella tenía curiosidad sobre lo que pasaba y eso, pero no reaccionó ni bien ni mal”• Si ha hablado con un ser querido, admite que éste le ha preguntado si ha cometido el delito

ANEXO M

COMPARACIÓN DE LAS PREDICCIONES DE LA BAI CON LAS PREDICCIONES DE LA (HMI) PARA CADA PREGUNTA DE LA ENTREVISTA. DATOS TOMADOS DE VRIJ, MANN Y FISHER (2006).

A NIVEL VERBAL

1 **¿Cuál crees que es el propósito de esta entrevista?**

Predicción *BAI*: quienes mientan serán más evasivos que quienes digan la verdad.

Predicción *HMI*: quienes mientan serán menos evasivos que quienes digan la verdad.

2 **¿Cometiste tú el delito?**

Predicción *BAI*: quienes mientan serán menos enfáticos en sus negaciones.

Predicción *HMI*: quienes mientan serán más enfáticos en sus negaciones.

3 **¿Sabes quién cometió el delito?**

Predicción *BAI*: es más probable que quienes mientan nieguen cualquier conocimiento.

Predicción *HMI*: no se establece ninguna predicción.

4 **¿Quién sospechas que podría haber cometido el delito?**

Predicción *BAI*: es menos probable que quienes mientan nombren a alguien.

Predicción *HMI*: no se establece ninguna predicción.

5 **¿Hay alguna otra persona distinta de ti de la que estés seguro que no ha cometido el delito?**

Predicción *BAI*: es menos probable que quienes mientan nombren a alguien que es inocente.

Predicción *HMI*: es más probable que quienes mientan nombren a alguien que es inocente.

6 **¿Crees que realmente alguien cometió el delito intencionadamente?**

Predicción *BAI*: es menos probable que quienes mientan admitan que realmente se haya cometido el delito intencionadamente.

Predicción *HMI*: es más probable que quienes mientan admitan que realmente se haya cometido el delito intencionadamente.

7 **¿Quién podría haber tenido la mejor oportunidad para cometer el delito?**

Predicción *BAI*: es menos probable que quienes mientan admitan que ellos hayan tenido alguna oportunidad.

Predicción *HMI*: no se establece ninguna predicción.

8 **¿Cómo te sientes al ser entrevistado respecto a la comisión del delito?**

Predicción *BAI*: es más probable que quienes mientan muestren sentimientos negativos.

Predicción *HMI*: no se establece ninguna predicción.

ANEXO M (CONTINUACIÓN)

9 ¿Has pensado alguna vez en realizar alguna acción similar al delito?

Predicción *BAI*: es más probable que quienes mientan admitan haber pensado alguna vez en cometer el delito.

Predicción *HMI*: no se establece ninguna predicción.

10 ¿Por qué crees que alguien pudo cometer el delito?

Predicción *BAI*: es menos probable que quienes mientan ofrezcan algún motivo razonable.

Predicción *HMI*: es más probable que quienes mientan ofrezcan algún motivo razonable.

11 ¿Qué crees que debería pasarle a la persona que cometió el delito?

Predicción *BAI*: es menos probable que quienes mientan sugieran un castigo serio.

Predicción *HMI*: es más probable que quienes mientan sugieran un castigo serio.

12 ¿Crees que hay circunstancias bajo las que la persona que cometió el delito debería ser perdonada?

Predicción *BAI*: es más probable que quienes mientan den una segunda oportunidad al culpable.

Predicción *HMI*: es menos probable que quienes mientan den una segunda oportunidad al culpable.

13 Dime que te habría impedido cometer el delito

Predicción *BAI*: es más probable que quienes mientan contesten en tercera persona.

Predicción *HMI*: no se establece ninguna predicción.

14 Una vez que finalice la investigación ¿Qué crees que indicarán los resultados respecto a tu implicación en el delito?

Predicción *BAI*: quienes mientan expresarán menos confianza en ser considerados inocentes.

Predicción *HMI*: no se establece ninguna predicción.

15 ¿A quién le has contado que ibas a tener esta entrevista hoy?

Predicción *BAI*: es menos probable que quienes mientan hayan contado algo sobre la entrevista a algún ser querido.

Predicción *HMI*: no se establece ninguna predicción.

A NIVEL NO VERBAL

Cruzar las piernas

Predicción *BAI*: más probable en quienes mientan.

Predicción *HMI*: menos probable en quienes mientan.

Cambiar de postura en la silla

Predicción *BAI*: más probable en quienes mientan.

Predicción *HMI*: menos probable en quienes mientan.

ANEXO M (CONTINUACIÓN)

Adaptadores (arreglarse el pelo o la ropa)

Predicción *BAI*: más probable en quienes mientan.

Predicción *HMI*: menos probable en quienes mientan.

Inclinarse hacia delante

Predicción *BAI*: más probable en quienes digan la verdad.

Predicción *HMI*: menos probable en quienes digan la verdad.

Establecer contacto ocular

Predicción *BAI*: más probable en quienes digan la verdad.

Predicción *HMI*: menos probable en quienes digan la verdad.

Ilustradores

Coinciden ambas predicciones

Predicción *BAI*: más probable en quienes digan la verdad.

Predicción *HMI*: más probable en quienes digan la verdad.

Respuestas rápidas

Coinciden ambas predicciones

Predicción *BAI*: más probable en quienes mientan.

Predicción *HMI*: más probable en quienes mientan.

Sinceridad

Predicción *BAI*: las respuestas de quienes mientan parecerán menos sinceras.

Predicción *HMI*: las respuestas de quienes mientan parecerán más sinceras.

Conductas resultantes de estados de ansiedad

Predicción *BAI*: más probable en quienes mientan.

Predicción *HMI*: menos probable en quienes mientan.

ANEXO N

FRASE Y DECLARACIÓN UTILIZADAS PARA EJEMPLIFICAR CADA CLAVE EN LOS VÍDEOS-EJEMPLO. LAS PARTES QUE APARECEN SUBRAYADAS REFLEJAN LOS PUNTOS EN LOS QUE SE INTRODUCIERON LAS CLAVES.

Frase:

“Entró y dejó sobre la mesa una carpeta negra”

Declaración completa:

“La mujer va caminando por un parque de una ciudad, al lado de una avenida. Entonces, llega a un banco al lado de una fuente de esas que tienes que apretar para que salga el agua, y se sienta. Deja las bolsas y el bolso que lleva a su lado en el banco y mira el reloj. Son unas bolsas como de comprar en una tienda de ropa, o sea, que no son las típicas bolsas de supermercado. Mira a su alrededor, parece como si buscara a alguien, o hubiera quedado con alguien a esa hora en ese sitio.

Luego, llega un hombre de traje, que parece un ejecutivo, con un periódico, y se sienta en el mismo banco que la mujer, pero no se conocen de nada, porque no se dicen nada. El hombre simplemente se sienta y se pone a leer el periódico. Al poco rato, llega andando una chica joven, de unos 25 años, va en chándal y lleva una pequeña mochila verde a la espalda, y en la mano lleva un plano. Al pasar al lado del banco donde está sentada la mujer, se agacha un poco para preguntarle cómo podía llegar a un sitio de la ciudad.

Entonces la mujer se levanta para señalarle en el mapa como podía llegar, de forma que se queda de espaldas a las bolsas de la compra y a su bolso, es decir, está al lado, pero de espaldas. Entonces, mientras le está explicando cómo llegar al sitio, el hombre del periódico empieza a meter las manos en el bolso de la mujer, y saca una cartera y un móvil, y se los guarda en su chaqueta, espera un poco y se va, mientras la mujer terminaba de explicar a la chica cómo llegar al sitio que le había preguntado. Luego la chica le da las gracias y se va, y la mujer se vuelve a sentar en el banco y sigue esperando sin enterarse de nada”.

ANEXO O

CUESTIONARIO DE TEMAS DE OPINIÓN.

A continuación se presentan una serie de temas de opinión sobre los que tienes que expresar tu postura. Para ello cada uno de los temas aparece junto con una escala cuyos extremos son -3 (totalmente en contra) y 3 (totalmente a favor). Para expresar tu opinión sobre cada tema, debes rodear con un círculo la cifra que mejor refleje lo que opinas. Además, debes puntuar en una escala de de -3 a 3 la importancia que tiene para ti cada uno de los temas sobre los que acabas de opinar, siendo -3 “no me importa en absoluto” y 3 “es muy importante para mí”.

Legalización del aborto

Totalmente En contra							Totalmente a favor							No me importa en absoluto					Es muy importante para mi		
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Ley anti-tabaco

Totalmente En contra							Totalmente a favor							No me importa en absoluto					Es muy importante para mi		
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Matrimonio entre homosexuales

Totalmente En contra							Totalmente a favor							No me importa en absoluto					Es muy importante para mi		
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Prohibición de fiestas de estudiantes en las facultades

Totalmente En contra							Totalmente a favor							No me importa en absoluto					Es muy importante para mi		
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Eutanasia

Totalmente En contra							Totalmente a favor							No me importa en absoluto					Es muy importante para mi		
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

ANEXO O (CONTINUACIÓN)

Creo que argumentar contra mis propias creencias está mal

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Traslado del archivo de Salamanca a Cataluña

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Legalización del hachís y la marihuana

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Investigación genética con células madre

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Aprobación del estatuto de Cataluña

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

Uso de la píldora del día después

Totalmente En contra					Totalmente a favor			No me importa en absoluto					Es muy importante para mi	
-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3	

ANEXO P

IMPRESO DE PARTICIPACIÓN VOLUNTARIA EN EL ESTUDIO.



UNIVERSIDAD DE SALAMANCA

FACULTAD DE PSICOLOGÍA

Departamento de Psicología Social y Antropología

Avda. de la Merced, 109-131
37005 Salamanca

La persona abajo firmante, D. D^a. _____,
con DNI _____, de sexo Varón / Mujer de _____ años de edad,
accede voluntariamente a participar en el experimento llevado a cabo en el
Departamento de Psicología Social y Antropología de la Universidad de Salamanca
_____, sin que su abandono antes de haber completado la tarea pueda
dar lugar a represalias de ningún tipo.

La persona abajo firmante no autoriza la divulgación de su identidad ni que,
explícitamente, se conecte su persona con los resultados de la investigación.

Finalmente, con independencia de si completa o no la tarea experimental, la persona
abajo firmante se compromete a no desvelar ni durante el experimento ni después de
éste, ante ninguna persona absolutamente ninguna información referente a su
participación en el presente experimento, hasta que reciba autorización expresa para
ello por parte de algún miembro del equipo investigador.

Salamanca, a _____ de _____ de 200____

Fdo.: _____

ANEXO Q

INSTRUCCIONES DEL CUESTIONARIO DE EVALUACIÓN PRE-ENTRENAMIENTO PARA EL GR-TRAD*.

A continuación vas a ver un vídeo en el que aparecen 10 personas. Cada una de estas personas hace una declaración de aproximadamente un minuto de duración. Cada declaración está basada en un vídeo que previamente ha visto el declarante. Después de haber visto el vídeo, cada declarante ha tenido 10 minutos para preparar su declaración sobre lo que aparecía.

A cada declarante, se le han hecho tres preguntas sobre lo que sucedía en el vídeo. En algunos casos veréis parte de una de las respuestas, en otros parte de dos y en otros, parte de las tres respuestas que cada declarante dio, por eso en algunos vídeos se puede apreciar un pequeño corte en la declaración. Esto es así porque quisimos que todos los fragmentos tuvieran la misma duración.

Cada una de las declaraciones puede o no ser mentira. Es decir, puede que el declarante haya mentado respecto a lo que haya visto en el vídeo o puede que no.

Tu tarea consiste en juzgar si cada declaración es mentira o no. Cada juicio lo emitirás al final de cada declaración, durante un tiempo en el que se detendrá el vídeo, rodeando con un círculo la palabra “No” o “Sí” para indicar si cada declarante miente. Además, en la escala de debajo, debes señalar, rodeando con un círculo, lo seguro que estas de cada uno de tus juicios, teniendo en cuenta que 1= totalmente inseguro y 7 = totalmente seguro.

Hay algunas **cosas que debes tener en cuenta** al efectuar tus juicios:

1. Había varios vídeos: No todos los declarantes vieron el mismo vídeo sobre el que tenían que hacer la declaración.
2. En la mayoría de ocasiones –aunque no siempre– los vídeos que vieron eran bastante parecidos entre ellos.
3. En la mayoría de ocasiones –aunque no siempre– los declarantes que mintieron no alteraron radicalmente lo que habían visto en el vídeo, sino que introdujeron pequeños cambios.

Esto hace que el contenido de las declaraciones a veces pueda ser parecido pero otras veces sea diferente, sin que ello se relacione con que sea mentira o no. Por esto, no debes emitir tus juicios comparando unas declaraciones con otras, sino juzgando cada declaración en sí misma, de forma absolutamente independiente del resto.

** Para el GR-ALT las instrucciones eran las mismas, excepto que en todos los casos en los que aparecía la palabra “mentira”, ésta se cambió por “verdad”. Para el GR-CONT se utilizaban ambas palabras. Es decir, se usaba la expresión “verdad o mentira” (por ejemplo, “Tu tarea consiste en juzgar si cada declaración es verdad o mentira”).*

ANEXO R

LOS SEIS REQUISITOS QUE DEBE CUMPLIR UN BUEN PROGRAMA DE ENTRENAMIENTO SEGÚN FRANK Y FEELEY (2003).

1. **Relevancia.** Se trata de adaptar o contextualizar el programa de entrenamiento a situaciones específicas que sean relevantes para el trabajo policial (por ejemplo interrogatorios). Esto es importante tanto para mantener el interés y la motivación de los profesionales como porque trabajar en contextos policiales supone analizar elementos conductuales similares a los que posteriormente los policías se encontraran realmente.
2. **Lo que está en juego.** Se trata de entrenar a los policías con material estimular (por ejemplo, vídeos de declaraciones) en el que lo que está en juego para la persona que hace la declaración es mucho. Además de simular mejor una situación real, en la que puede estar en juego, por ejemplo, la libertad de una persona, en este tipo de situaciones se distingue mejor cuándo alguien dice la verdad y cuándo miente.
3. **Entrenamiento.** Se trata de diseñar un protocolo que transmita el conocimiento clara y efectivamente. En este punto, Frank y Feeley hacen hincapié en que un programa de entrenamiento que incluya información teórica, fases prácticas y *feed-back* debe administrarse en sesiones de alrededor de 50 minutos a lo largo de varios días, hasta cerciorarse de que las personas entrenadas han adquirido el conocimiento que se trata de transmitir.
4. **Evaluación.** En este punto, los autores hacen referencia a la necesidad de contar con adecuadas pruebas pre- y post-entrenamiento que servirán para observar en qué medida el entrenamiento ha sido efectivo.
5. **Capacidad de generalizar a otras situaciones.** Se trata de desarrollar un entrenamiento que sea útil en el mayor número posible de situaciones. Para ello, dado que no existen conductas que aparezcan de forma indefectible siempre que una persona miente o siempre que dice la verdad, han de seleccionar adecuadamente aquellos indicios que la investigación científica haya encontrado más fiables.
6. **Generalidad en el tiempo.** El entrenamiento debe ser tal que los beneficios conseguidos a partir de él deben perdurar en el tiempo. Es decir, se trata de crear un *modus operandi* estable, de forma que el conocimiento y la forma de proceder adquiridos se sigan utilizando una vez que el entrenamiento haya finalizado. En este punto, Frank y Feeley plantean que los efectos perniciosos de los estereotipos y las falsas creencias son tan fuertes que pueden reaparecer después de cierto tiempo, con lo que indican la conveniencia de que los profesionales se sometan a sesiones posteriores de entrenamiento a lo largo de toda su carrera profesional.