



Acción Psicológica

ISSN: 1578-908X

accionpsicologica@psi.uned.es

Universidad Nacional de Educación a

Distancia

España

ARCE, RAMÓN

ANÁLISIS DE CONTENIDO DE LAS DECLARACIONES DE TESTIGOS: EVALUACIÓN
DE LA VALIDEZ CIENTÍFICA Y JUDICIAL DE LA HIPÓTESIS Y LA PRUEBA FORENSE

Acción Psicológica, vol. 14, núm. 2, diciembre-, 2017, pp. 171-190

Universidad Nacional de Educación a Distancia

Madrid, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=344054646012>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

ANÁLISIS DE CONTENIDO DE LAS DECLARACIONES DE TESTIGOS: EVALUACIÓN DE LA VALIDEZ CIENTÍFICA Y JUDICIAL DE LA HIPÓTESIS Y LA PRUEBA FORENSE

CONTENT ANALYSIS OF THE WITNESS STATEMENTS: EVALUATION OF THE SCIENTIFIC AND JUDICIAL VALIDITY OF THE HYPOTHESIS AND THE FORENSIC PROOF

RAMÓN ARCE

Cómo referenciar este artículo/How to reference this article:

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content Analysis of the Witness Statements: Evaluation of the Scientific and Judicial Validity of the Hypothesis and the Forensic Proof]. *Acción Psicológica*, 14(2), 171-190. <https://doi.org/10.5944/ap.14.1.21347>

Resumen

La valoración de la credibilidad del testimonio es el elemento central en la gran mayoría de las decisiones judiciales, especialmente cuando se enfrentan versiones contradictorias de los hechos. La evaluación del testimonio basada en indicios cognitivos (verbales), específicamente la basada en la hipótesis Undeutsch que sostiene que la memoria de un evento auto-experimentado difiere en contenido y calidad (definida por criterios de realidad) de la memoria de un evento no experimentado, se ha mos-

trado como la más efectiva en la valoración de la credibilidad. De esta hipótesis se derivó una técnica forense, el Statement Reality Analysis (SRA), que incluye un listado de criterios de realidad, el Criteria Based Content Analysis (CBCA). Revisiones meta-analíticas prestan apoyo a la capacidad discriminativa entre memorias de eventos auto-experimentados y no experimentados de las categorías de realidad del CBCA, así como de la puntuación total del CBCA, en todo tipo de poblaciones (menores y adultos) y testigos (denunciantes, testigos y denunciados), y de memorias (abusos sexuales, violencia de género). Una hipótesis para ser admitida como científica

Correspondencia: Ramón Arce. Facultade de Psicoloxía, Universidade de Santiago de Compostela (España).

Email: ramon.arce@usc.es

ORCID: Ramón Arce (<http://orcid.org/0000-0002-5622-3022>)

Universidade de Santiago de Compostela, España.

Recibido: 7 julio de 2017.

Aceptado: 28 de septiembre de 2017.

en la Sala de Justicia ha de cumplir los criterios Daubert que, constatamos, cumple la hipótesis Undeutsch. No obstante, este soporte de la evidencia científica a la hipótesis a través de los criterios de realidad no significa que la técnica forense sea válida. Revisada la técnica verificamos que no cumple los criterios científicos (el CBCA no es un sistema metódico, esto es, fiable y válido; y adolece de validez criterial al no incluir criterio de decisión objetivo y estricto) y legales y jurisprudenciales (no garantiza el cumplimiento del principio constitucional de presunción de inocencia; la entrevista de obtención de la declaración puede dar lugar a una prueba carente de validez jurídica; no evalúa adecuadamente la persistencia en la incriminación). Finalmente, se describe una técnica forense, el Sistema de Evaluación Global (SEG), basada en análisis de contenido de las declaraciones que solventa estas limitaciones del SVA.

Palabras clave: Statement Validity Analysis; Criteria Based Content Analysis; Sistema de Evaluación Global; Credibilidad del testimonio; Admisión de la prueba.

Abstract

Assessing the credibility of a testimony is a crucial step in judicial decision-making, primarily when inconsistencies arise due to conflicting versions of events. The evaluation of the veracity of a testimony based on verbal cues, in particular those based on the Undeutsch hypothesis that contends that the memory of self-experienced real-life events differ in content and quality (as defined by reality criteria) from the memory of fabricated or fictitious accounts, is considered to be the most effective tool for assessing credibility. This hypothesis has given rise to a forensic technique known as Statement Validity Analysis (SVA) that includes a list of reality criteria i.e., Criteria Based Content Analysis (CBCA). Meta-analytical reviews support the discriminatory power of CBCA reality criteria and the CBCA total score in distinguishing between the memory of self-experienced real-life events versus fabricated or fictitious accounts in all types of populations (children and adults), witnesses (plaintiffs, witnesses, and defendants), and memories (sexual abuse, gender

violence). A hypothesis must fulfil the Daubert standards for it to be admitted as scientific evidence in a court of law, as is the case of the Undeutsch hypothesis. Notwithstanding, the scientific evidence upholding this hypothesis in terms of reality criteria does not substantiate the validity of this forensic technique. A review of this technique revealed it failed to meet scientific criteria (CBCA is not a methodic system, and is neither valid nor reliable; and it lacks criterion validity as there is no strict objective decision criterion); and has legal and jurisprudential flaws (no safeguards guaranteeing the constitutional principle of the presumption of innocence; witness interviews to obtain statements may lack judicial validity; and fail to evaluate appropriately the consistency of the evidence). Finally, a forensic technique based on the content analysis of statements, the Global Evaluation System (GES), is examined in order to overcome the limitations of SVA.

Keywords: Statement Validity Analysis; Criteria Based Content Analysis; Global de Evaluation System; Testimony credibility; Proof admissibility.

Introducción

La evaluación de la veracidad del testimonio es la piedra angular sobre la que pivotan las decisiones judiciales (Novo y Seijo, 2010), especialmente en los delitos cometidos en la esfera privada, tales como los delitos sexuales y de violencia contra las personas (e.g., violencia familiar y de género). Al respecto, Hans y Vidmar (1986) han estimado que esta contingencia afecta aproximadamente al 85 % de los juicios a nivel mundial. Como quiera que la carga de la prueba recae en la acusación, corresponde a ésta demostrar la autenticidad de los hechos denunciados, pudiendo ser suficiente, acorde a la jurisprudencia, el testimonio del o la denunciante, al ser prueba directa y no indiciaria. Del interés que todo denunciante tiene en la condena del acusado, no se desprende directamente que su testimonio sea falso, pudiendo el testimonio del testigo-víctima ser suficiente para enervar la presunción de inocencia (Sentencia 210/2014 del TS, Sala de lo Penal, 14 de marzo de 2014). No obstante, si concurre o concurriera en el denunciante algún

móvil tales como resentimiento, enemistad, venganza, enfrentamiento, interés o cualesquiera otros de índole parecida (principio de incredibilidad subjetiva), éste priva al testimonio del denunciante de la aptitud probatoria necesaria para otorgar certidumbre, generando en el Juzgador una duda razonable sobre la veracidad de los hechos relatados por el denunciante (no enerva el principio de presunción de inocencia). La deficiencia en este principio de evaluación judicial de la credibilidad del testimonio, muy frecuente en casos de violencia de género y familiar, no invalida la declaración, puede compensarse con un reforzamiento de los principios de verosimilitud y persistencia en la incriminación. Específicamente, el que el testimonio esté apoyado por corroboraciones periféricas (principio de verosimilitud) que validen el testimonio del denunciante, suele compensar la debilidad en la credibilidad subjetiva, enervando la presunción de inocencia. Dentro de éstas, la prueba psicológica forense sobre la honestidad del testimonio (para ello, se han usado indistintamente, en la literatura psicológica, términos como verosimilitud, realidad o credibilidad. No obstante, en el contexto judicial no ha de utilizarse el término credibilidad para la prueba psicológica, pues éste es un concepto jurídico, no correspondiendo por ello al psicólogo informar sobre la credibilidad. Además, la prueba psicológica no evalúa credibilidad, sino realidad, atributos de memoria o criterios de contenido de memoria), limitándose a ser un medio (prueba indiciaria) para atribuir credibilidad objetiva al testimonio del denunciante. En esta línea, un análisis de sentencias judiciales en España halló que en el 93.3 % de las periciales psicológicas que dotaban de verosimilitud al testimonio de testigos víctimas (incluye todo tipo de poblaciones y delitos), la sentencia era condenatoria, en tanto que cuando no se la conferían, la sentencia fue absolutoria en el 100 % de los casos (Novo y Seijo, 2010). Así pues, el informe de credibilidad del testimonio es suficiente para enervar el principio de presunción de inocencia ante la alegación o concurrencia de potenciales móviles (incredibilidad subjetiva) que generan en el juzgador incertidumbre sobre la sinceridad del testigo. Por último, el relato en el que se sustenta la acusación ha de ser persistente (principio de persistencia en la incriminación), es decir, un relato coherente (i.e., conexión entre las partes del evento), sin modificaciones importantes en el tiempo del relato (i.e., desdecirse, contradecirse), y concreto (i.e., sin ambigüedades, generali-

dades, vaguedad). En algunos tipos de casos cometidos en la esfera privada, especialmente agresiones sexuales y violencia familiar y de género, con cierta frecuencia, las/los denunciantes modifican en el tiempo sustancialmente el relato (e.g., se desdicen, retiran la denuncia, se autoacusan de denuncia falsa, cambian elementos centrales del relato), o no cooperan con la justicia en la prestación del testimonio (e.g., relatos vagos, ambiguos, generales, incoherentes, ocultan información, se acogen al derecho a no declarar contra un familiar). Bajo esta contingencia, también, la pericial psicológica de la credibilidad del testimonio vuelve, de nuevo, a ser central pues, obtenido un relato persistente en dos momentos, puede justificar la falta de persistencia con otras declaraciones policiales, judiciales o ante testigos, atribuyendo la falta de consistencia a la técnica de entrevista o a un daño consecuencia de la victimización. No obstante, esta alta tasa de aceptación ha cambiado con el paso del tiempo. Así, aunque no sólo está referida a la credibilidad del testimonio, sino también daño psicológico, Arce (2015) con los datos del Observatorio contra la Violencia Doméstica y de Género del Consejo General del Poder Judicial, el Ministerio de Igualdad, y el Ministerio de Sanidad, Servicios Sociales e Igualdad de España encontró que la tasa de condena era de $\pm 22\%$, siendo la causa fundamentalmente la falta de prueba que otorgara credibilidad al testimonio de la denunciante. Dado que las declaraciones de la denunciante y denunciando suelen ser contradictorias, si no se puede afirmar como verdadera la de la denunciante ha de aceptarse la más beneficiosa al acusado (Sentencia 246/2016 de la AP de Las Palmas de Gran Canaria, Sala de lo Penal, de 28 de junio de 2016), para que la declaración de la denunciante adquiera mayor verosimilitud, jueces y tribunales han de recurrir a los conocimientos provenientes de la psicología del testimonio (Boente, 2011). De acuerdo con la *Guía y Manual de Valoración Integral Forense de la Violencia de Género y Doméstica* (Ministerio de Justicia, 2005), practicada desde los Institutos de Medicina Legal (corresponde a la acusación demostrar la culpabilidad del investigado y ésta se vale de los forenses de dichos institutos), se incluye en el plan de actuación la valoración de la credibilidad del testimonio de la agredida/denunciante que realiza el psicólogo, siendo la técnica psicológica de referencia el Statement Validity Analysis (SVA), que define los pasos a dar en la implementación de la técnica (ver

Tabla 3), criterios de Validez de la Declaración (ver Tabla 5), y criterios de contenido de la declaración, Criteria Based Content Analysis (CBCA) (ver Tabla 4), para estimar la credibilidad del testimonio basado en los criterios de realidad.

Evaluación de cumplimiento de la hipótesis Undeutsch de los criterios Daubert

Muchos han sido los procedimientos a los que se ha acudido históricamente para la corroboración periférica (prueba indiciaria) de la credibilidad del testimonio: correlatos del comunicador, indicios no verbales y paraverbales, indicios psicofisiológicos e indicios cognitivos (Arce y Fariña, 2013b; Vrij, 2008). De todos ellos, los indicios cognitivos (verbales) no sólo son los que han arrojado mejores resultados en la clasificación de testimonios verídicos, sino que gozan de un alto valor judicial probatorio en países como España, Alemania, Holanda, Suecia o Estados de EE.UU. (Novo y Seijo, 2010; Steller y Böhm, 2006). Aunque los indicios cognitivos se sustentan desde más de una explicación teórica, la conocida como *hipótesis Undeutsch* es la más manida. Esta sostiene que la memoria de un evento real (auto-experimentado) difiere en contenido y calidad (que ha de ser definida en término de criterios de realidad) de una memoria de un evento no experimentado (e.g., inventado, fabricado, ficticio, tergiversado, fantasioso) (Undeutsch, 1967, p. 125). En dicha hipótesis descansa la técnica de evaluación forense de la realidad del testimonio, basada en criterios de realidad (indicios cognitivos), el SVA, que se ha mostrado como una herramienta potente y muy eficaz en la discriminación en todo tipo de poblaciones (menores y adultos) y testigos (denunciantes, testigos y denunciados), y en todo tipo de memorias (abusos sexuales, violencia de género,...), entre memorias de hechos auto-experimentados/vividos y fabricados (Amado, Arce y Fariña, 2015; Amado, Arce, Fariña y Vilariño, 2016).

El Tribunal Supremo de EE.UU. en el caso *Daubert v. Merrell Dow Pharmaceuticals* (1993) estableció los criterios que debe cumplir una prueba para ser admitida como científica en la Sala de Justicia y que se asumen científicamente como propios: 1) ¿Puede la teoría (hipó-

tesis, en nuestro caso) o técnica ser contrastada? No sólo es posible, sino que ha sido sometida a prueba tanto en estudios experimentales (de laboratorio) como de campo. Cada uno de estos diseños tiene sus ventajas e inconvenientes. Los estudios experimentales presentan como ventaja un alto grado de control de las variables a manipular de modo que la fiabilidad de la clasificación de los relatos como reales o fabricados es total, en detrimento de la validez externa, especialmente en el campo de la investigación en psicología del testimonio. A este respecto y en este ámbito, la validez externa de este tipo de estudios ha sido etiquetada como validez aparente (Konecni y Ebbesen, 1992). Para afrontar esta contingencia se han perfeccionado los diseños experimentales creando condiciones que se acerquen a la realidad, en lo que se han denominado diseños de alta fidelidad. En todo caso, no dejan de ser aproximaciones distantes de la realidad porque, por ejemplo, en los abusos sexuales a niños no se pueden recrear, por obvias razones éticas, memorias inventadas de niños objeto de tales acciones. De hecho, en un estudio que comparaba los resultados del testimonio de una investigación experimental de alta fidelidad y un experimento de campo mostraron resultados diferentes (Fariña, Real y Arce, 1992). Todo esto no pasó inadvertido para el tribunal que concretó que no era suficiente que la teoría fuera contrastada sólo en estudios de laboratorio, sino también en estudios de campo. Ahora bien, sirven de preliminares a los de campo y, entre ambos, pueden validarse.

Por su parte, los estudios de campo tienen como fortaleza la validez externa de los diseños y como debilidad la validez interna, al no tener un control sobre las variables a manipular, esto es, sobre el criterio de clasificación (ground truth) de los eventos reales. La investigación en este contexto ha tomado criterios de clasificación tan dispares como la sentencia judicial, la confesión del acusado, evidencia médica o registros poligráficos. La falta de fiabilidad de estos criterios (mayores en unos que en otros) es el asunto clave. Sucintamente, de las nueve amenazas a la validez interna que relaciona Campbell (1969) de los estudios de campo, la realmente afectada en este contexto de investigación es la *selección* de los casos. Si la selección fuera improcedente afectaría a los resultados. Pues bien, la selección de los casos fabricados para la comparación con los casos de campo es igual a la

de los estudios experimentales, por lo que sólo tendría efectos en los casos reales clasificados por los ground truth. Además, mientras en los estudios experimentales los efectos se han aislado de modo que sólo pueden atribuirse a la manipulación experimental, en los de campo, al no manipular el investigador la variable independiente o causal, pueden ser consecuencia de otras causas (hipótesis alternativas). Si bien, la existencia de hipótesis alternativas a la experimental, si la confirman los resultados, no es suficiente para invalidar ésta (como máxima derivada de los estudios experimentales, sólo se invalida cuando las alternativas son más fiables y plausibles), esta potencial fuente de error en el diseño se puede controlar con ciertas salvaguardas. Primera, si la hipótesis también es confirmada por los estudios experimentales gozaría de validez convergente. En otras palabras, sería robusta a las fuentes de error asociadas a los diseños de los estudios experimentales (validez externa) y de los estudios de campo (validez interna). Segunda, si la hipótesis se confirma en diferentes contextos (moderadores que podrían mediar efectos en la misma), entonces dispondrá de consistencia inter-contextos. Tercera, si la hipótesis no es rechazada por los estudios (resultados inconvenientes, esto es, tamaños del efecto promedio negativos/contrarios a la hipótesis), se obtiene validez convergente. Cuarta, si el factor ground truth no media diferencias en los resultados, se deduce que hay consistencia inter-criterios, validándose unos a otros. 2) ¿Ha sido probada la técnica o hipótesis? No sólo se puede someter a prueba, sino que ha sido probada en muchos estudios, mostrándose válida (i.e., estadísticamente significativa en la discriminación entre memorias de hechos experimentados y fabricados) tanto en estudios experimentales ($\delta = 0.56$ y 0.32 , en población de menores y adultos, respectivamente) y como de campo ($\delta = 2.71$ y 0.45 ; en menores y adultos); en testimonios de denunciantes en poblaciones de adultos ($\delta = 0.56$) y de menores ($\delta = 0.79$); testigos presenciales ($\delta = 0.51$) y encausados ($\delta = 0.35$); y en diversos tipos de delitos (e.g., abuso sexual infantil, $\delta = 0.79$; agresión sexual y de género, $\delta = 0.87$) y contextos (memorias de hechos delictivos y no delictivos; memorias de hechos fabricadas, inducidas o sugeridas) (Amado et al., 2015, 2016). Al hallarse apoyo en todos estos ámbitos se puede concluir que la hipótesis es robusta, es decir, no varía en las diferentes condiciones de estudio. 3) ¿Se conoce la tasa de error? Bien,

como quiera que la hipótesis fue formulada y no demostrada por el autor y que los sistemas categoriales se crearon con base en la experiencia de los autores (top-down) obtenida de miles de casos, y sin un criterio de decisión objetivo y estricto, no la pusieron a prueba, por lo que no informaron de la tasa de error. Sin embargo, en los meta-análisis de Amado et al. (2015, 2016) se han calculado estadísticamente los márgenes de error, tras un estudio sistemático de la literatura. Éstos se cifran en el 31.5 % de falsos negativos en la población de menores y del 36.5 % en adultos. 4) ¿Ha sido sometida la hipótesis/técnica a un proceso de revisión por pares y publicada (no es una condición sine qua non)? En más de 800 ocasiones (tamaños del efecto) (Amado et al., 2015, 2016). En consecuencia, los resultados sobre la validez de la hipótesis son, en terminología popperiana (Popper, 1961), un *hecho*, esto es, son altamente fiables. 5) ¿Es la teoría científica o la técnica aceptada generalmente como válida por la comunidad científica? El tribunal cita como un medio la consulta a la comunidad científica sobre el método y principios de la hipótesis. Una aceptación amplia le conferiría un valor de prueba admisible –no es un prerequisite absoluto la aceptación general–, en tanto que una aceptación baja, debería conducir a tomar la prueba con escepticismo. No hemos encontrado tal consulta ni de esta hipótesis, ni en ninguna otra psicológica. Desde una óptica científica, un medio de consulta a la comunidad es la revisión (sistemática, meta-análisis) de la literatura de modo que, si ésta le presta apoyo general a la hipótesis, esto quiere decir que la comunidad científica la valida (aceptación general). Los referidos meta-análisis no sólo le prestan apoyo general, sino que no hay ningún resultado meta-analítico que la contradiga (los resultados meta-analíticos confirman un valor discriminativo significativo entre memorias del total de criterios de realidad, CBCA, esto es, confirmación general de la hipótesis).

Evaluación de la técnica de análisis de contenido de las declaraciones SVA/CBCA en los criterios científicos, y legales y jurisprudenciales

Ahora bien, Arce (2017) advirtió que el cumplimiento de la hipótesis de los criterios Daubert no presupone

directamente que lo haga la técnica derivada de ésta (falacia ecológica), al tiempo que puso de manifiesto que los criterios relacionados por la Corte Suprema de EE.UU. no satisfacían las demandas que se han de requerir a una prueba forense (técnica), concretando un listado de criterios científicos adicionales (ver Tabla 1) y legales y jurisprudenciales (ver Tabla 2).

Tabla 1

Criterios científicos

El instrumento de medida ha de ser fiable y válido.
La técnica subyacente debe ser falseable, refutable y sometible a prueba y replicable.

- La aplicación de la técnica ha de permitir la revisión externa.
- Se han de poder comprobar los métodos usados en la aplicación de la técnica.

Se ha de estimar la aplicación de la técnica al caso en cuestión.
Validez criterial: La técnica ha de incluir un criterio de decisión objetivo y estricto.

Tabla 2

Criterios legales y jurisprudenciales

¿Ha sido aplicada la técnica por un perito titular?
¿Garantiza la técnica el cumplimiento del principio de presunción de inocencia?
¿Cumple el método de obtención de la prueba con los estándares legales y jurisprudenciales?
¿Evalúa la técnica la persistencia?
¿Puede ser la aplicación de la técnica revisada por otros peritos?
¿Puede ser replicada la prueba por otro perito?

De las técnicas forenses resultantes de la hipótesis, el SVA/CBCA (Steller y Köhnken, 1989), se consigna co-

mo la técnica de referencia. No en vano, resulta de una integración de las anteriores (Arntzen, 1970; Dettenborn, Froehlich, y Szewczyk, 1984; Szewczyk, 1973; Undeutsch, 1967). El SVA describe el procedimiento (ver Tabla 3) y criterios de validez de la prueba (ver Tabla 4), en tanto el CBCA los criterios de realidad (ver Tabla 5).

Tabla 3

Procedimiento en la aplicación del SVA

1. Estudio del caso.
2. Entrevista semi-estructurada del testigo.
3. Análisis de contenido de la declaración basado en criterios, CBCA (ver Tabla 4).
4. Listado de criterios de validez (ver Tabla 4).

Tabla 4

Listado de validez del SVA

- a) Características psicológicas:
 - (In)adecuación del lenguaje y conocimientos.
 - (In)adecuación del afecto.
 - Susceptibilidad a la sugestión.
- b) Características de la entrevista:
 - Entrevista coercitiva, sugestiva o dirigida.
 - (In)adecuación global de la entrevista.
- c) Motivación:
 - Motivos para la denuncia.
 - Contexto en que se produjo la denuncia o revelación original.
 - Presiones para presentar una denuncia falsa.
- d) Cuestiones de la investigación:
 - (In)consistencia con las leyes de la naturaleza.
 - (In)consistencia con otras declaraciones.
 - (In)consistencia con otras pruebas.

Tabla 5

Criterios de realidad del CBCA

-
- | | |
|----|--|
| a) | Características generales: |
| | <ul style="list-style-type: none"> - Estructura lógica (coherencia y consistencia interna). - Elaboración inestructurada (presentación desorganizada). - Cantidad de detalles (abundancia de detalles o hechos distintos). |
| b) | Contenidos específicos: |
| | <ul style="list-style-type: none"> - Engranaje contextual (ubicación de la narración en un espacio y tiempo). - Descripción de interacciones (cadena de acciones entre el testigo y otros actores). - Reproducción de conversación (réplica de conversaciones). - Complicaciones inesperadas durante el incidente (por ejemplo, interrupción imprevista). |
| c) | Peculiaridades del contenido: |
| | <ul style="list-style-type: none"> - Detalles inusuales (detalles con baja probabilidad de ocurrencia). - Detalles superfluos (detalles irrelevantes que no contribuyen significativamente a los hechos). - Incomprensión de detalles relatados con precisión (explicitación de detalles que el menor no comprende, pero realmente sí tienen sentido). - Asociaciones externas relacionadas (inclusión de información externa a los hechos en sí pero relacionada con ellos, tal como en una agresión sexual recordar conversaciones anteriores sobre este tema). - Relatos del estado mental subjetivo (referencias a sentimientos, emociones o cogniciones propias). - Atribución del estado mental del autor del delito (referencias al estado mental del agresor y atribución de motivos). |
| d) | Contenidos referentes a la motivación: |
| | <ul style="list-style-type: none"> - Correcciones espontáneas (correcciones espontáneas o mejoras de la declaración). - Admisión de falta de memoria (reconocimiento de lagunas de memoria). - Plantear dudas sobre el propio testimonio. - Auto-desaprobación (actitud crítica sobre su propia conducta). - Perdón al autor del delito (la declaración de la víctima favorece al acusado, o evitación de más acusaciones). |
| e) | Elementos específicos de la agresión: |
| | <ul style="list-style-type: none"> - Detalles característicos de la ofensa (descripciones que contradicen las creencias habituales sobre el delito). |
-

Conocido que las categorías de contenido pueden discriminar entre memorias de hechos vividos y fabricados, y, lo que es más relevante para la práctica forense, clasificar correctamente la memoria de hechos vividos, no implica que dichos sistemas sean metódicos, es decir, fiables y válidos (Weick, 1977). La fiabilidad de las categorías (y, por extensión, de los sistemas como el global de todas las categorías; véase Amado et al., 2015, para el cómputo de la fiabilidad del sistema) se ha medido a través del acuerdo entre codificadores (la concordancia no es fiabilidad) o la correlación (fiabilidad) entre codificaciones en la codificación de las categorías de análisis. En todo caso, estas medidas no son de la fiabilidad del sistema categorial, sino de la medida, es decir, del acuerdo o consistencia en la codificación entre codificadores. Además, estas medidas del acuerdo/consistencia intercodificadores están sujetas a muchas fuentes de error: (a) porque suele ser entre dos codificadores lo que no garantiza la consistencia o acuerdo con otros codificadores (la

consistencia/fiabilidad verdadera ha de estimar la misma con cualquier otro codificador que aplique la misma categoría o sistema); (b) en un momento dado, lo que no garantiza la persistencia de la medida en el tiempo (test-retest); y (c) una medida directa (recuento de las contingencias de cada categoría por cada codificador; clasificación de la categoría como presente o ausente; o grado de presencia) de las categorías de contenido (registro) tampoco garantiza tal (esto es, una categoría codificada como presente por los evaluadores no quiere decir que realmente sean consistentes. De hecho, si se refieren a contingencias diferentes en la codificación de la categoría, serían errores de codificación y no aciertos. A este tipo de error lo denominamos error sistemático de evaluación de la codificación y, para afrontarlo, creamos un procedimiento de control de esta fuente de error, consistente en la verificación de la correspondencia entre las codificaciones; Fariña, Arce y Novo, 2002). Como no se ha estimado la fiabilidad del sistema, tampoco la validez

del mismo (CBCA) en sí, dado que fiabilidad y validez se relacionan. En otras palabras, la validez discriminante de las categorías valida la hipótesis, pero no así el instrumento (sistema categorial). Adicionalmente, el sistema categorial, CBCA, incumple otros parámetros para que el instrumento sea válido (Bardin, 1996): las categorías de análisis no son mutuamente excluyentes (Horowitz et al., 1997; Roma, San Martini, Sabatello, Tatarelli y Ferracuti, 2011), lo que conlleva a duplicidad de medidas; esto es, que se registren categorías de modo duplicado, tal que da entrada a un error de sobre-estimación de la realidad; homogeneidad (esto es, consistencia interna del sistema que es desconocida); exhaustividad (las categorías de realidad del CBA no son exhaustivas pudiendo añadirse más de modo que la eficacia del sistema está mermada; Steller y Köhnken, 1989; Amado et al., 2016; Arce y Fariña, 2009); fidelidad (el SVA no contiene una herramienta para el control de la fidelidad de la aplicación de la técnica en la práctica forense); objetividad (las categorías no se han definido con precisión dando lugar a interpretaciones y codificaciones diversas, no se ha identificado la unidad de análisis, ni la medida de cada categoría); pertinencia (las categorías de contenido del CBCA fueron construidas para memorias de niños –el concepto niños aquí no es tal pues abarcaba legalmente, cuando se formuló la hipótesis (Undeutsch, 1989), hasta los 21 años y la aplicación de la técnica se ha generalizado hasta los 18 por lo que realmente no se trata de niños– y abusos sexuales, encontrándose que eran pertinentes, pero no generalizables (Amado et al., 2016); y la hipótesis es para todo tipo de casos, en tanto el engaño interactúa con el contexto; Volbert y Steller, 2014); y productividad (parámetro que cumple, dado que todos los criterios de realidad del CBCA resultaron productivos y significativos en la discriminación entre memorias de eventos vividos y fabricados). En el SVA no se concreta un procedimiento que facilite la revisión externa y los métodos usados (e.g., si la entrevista ha introducido preguntas engañosas o sugestivas), pero se podría aplicar sin mayor problema. Bastaría con introducir la salvaguarda de que las entrevistas y codificación han de quedar registradas para que pueda contrastarse si la técnica y el método (e.g., los registros de las categorías; entrevista) han sido aplicados correctamente (en la práctica forense, por las demandas judiciales, ya se ha convertido en norma el registro y almacenamiento en vídeo de las entrevistas).

Asimismo, el SVA nada dice de si la prueba sería replicable, pero ciertamente lo es. Las entrevistas narrativas en recuerdo libre pueden obtenerse en series repetidas sin que contaminen el contenido de la declaración (Memon, Wark, Bull y Köhnken, 1997). Tampoco incluye la técnica una estimación de la aplicación de la misma al caso. En otras palabras, el SVA no contiene herramienta para conocer si el perito ha aplicado correctamente la técnica al caso en cuestión (fiabilidad de la medida en diseños de $N = 1$). El último de los criterios científicos, la validez criterial, tampoco la cumple el SVA de un modo satisfactorio; de facto, no concreta un criterio de decisión objetivo y estricto como se le requiere a una prueba forense. Las propuestas de criterio de decisión han sido varias, pero ninguna un criterio objetivo y estricto. Así, Steller y Köhnken (1989) calificaron el criterio de decisión de semi-estandarizado, en tanto Undeutsch (1989) de semi-objetivo. Las propuestas, todas ellas cualitativas, han ido desde la mención a reglas de evaluación (intensidad de manifestación del criterio, número de manifestaciones del criterio, capacidad para declarar del testigo y características del evento; Undeutsch, 1989); a una escala de 5 puntos (creíble, probablemente creíble, indeterminado, probablemente no creíble, increíble; Steller, 1989); a estimaciones clínico-intuitivas (Steller y Köhnken, 1989); hasta un juicio clínico (Köhnken, 2004). En suma, el SVA no cumple con la demanda de objetividad en la conclusión que se requiere a una técnica forense.

En lo relativo a los criterios legales y jurisprudenciales, la Ley de Enjuiciamiento Criminal española discierne entre *peritos titulares*, aquellos que tienen título oficial de una ciencia que esté reglamentado por la administración y *peritos no titulares*, sin dicho título (Art. 457), debiendo valerse el juez, jueza o tribunal con preferencia de los primeros (Art. 458). En el contexto de capacitación profesional actual, serían peritos titulares aquellos con un título de Máster Oficial que contenga en las competencias profesionales las de perito en esta materia. En la formulación original del CBCA, Steller y Köhnken (1989), además de a psicólogos, atribuyen dicha competencia a policías, trabajadores sociales y otros profesionales del entorno judicial entrenados en la técnica. Bien, este requerimiento legal puede (y ha de) corregirse en la dirección de cumplir con adjudicar la aplicación de la técnica a peritos titulares.

Dado que las técnicas de análisis de contenido de la declaración basadas en criterios se aplican como prueba de cargo, la conclusión ha de adherirse al principio constitucional de presunción de inocencia (art. 24.2º CE) que se concreta en la máxima “es esencial que los inocentes estén en todo caso protegidos frente a condenas infundadas, mientras que es suficiente que los culpables sean generalmente castigados” (Sentencia 213/2002 del TS, Sala de lo Penal, 14 de febrero de 2002). En consecuencia, la prueba forense indiciaria de análisis de contenido del testimonio no ha de incurrir en clasificar un testimonio no real como real, ya que dimanaría en la condena de un inocente. A este respecto, el SVA no contiene herramienta o criterio de decisión que controle la incorrecta clasificación de un evento fabricado como real, y, en consecuencia, no han calculado dicho error. Los cálculos de las probabilidades de falsos positivos (clasificación de casos fabricados como verdaderos) con la puntuación total del CBCA derivados de los meta-análisis de Amado et al. (2015, 2016) son como promedio de .315 para menores y .365 para adultos. Es decir, el SVA no garantiza el cumplimiento del principio constitucional de presunción de inocencia.

La prueba, además, ha de ser obtenida lícitamente (constitucionalmente; art. 24 de la CE) y de modo que tenga validez jurídica. Por ello, no se puede practicar sobre el encausado como prueba de cargo, ya que éste está protegido por el derecho a no declarar contra sí mismo, a estar asistido por un letrado en la declaración y a no confesarse culpables (decir verdad); esto es, la prueba de ser constitucionalmente obtenida. Adicionalmente y por analogía con las pruebas biológicas, la prueba de análisis de contenido del testimonio ha de obtenerse de modo que no esté contaminada. La fuente de contaminación en ésta dimana de la entrevista, tal que si el entrevistador contaminara la prueba ésta no tendría validez jurídica. Las entrevistas en el ámbito judicial se clasifican en interrogativas y narrativas, no siendo válidas para la práctica forense las entrevistas puramente interrogativas dado que contaminan la prueba (e.g., sugestión, coacción, inducción) y no propician la manifestación de criterios de contenido. Por su parte, las narrativas dan lugar a una memoria amplia del evento, necesaria para la práctica de la prueba que, asimismo, facilita la recuperación de mayor cantidad de información sobre el evento, la discrimi-

nación entre declaraciones de hechos vividos y fabricadas, y la aparición de criterios de contenido (Sooniste, Granhag, Strömwall y Vrij, 2015; Vrij, 2005). La entrevista recogida en el SVA es de tipo narrativo seguida de interrogatorio en el que, advierte, se han de evitar preguntas coercitivas, sugestivas o dirigidas, al tiempo que ha de ser adecuada (e.g., resaltar al entrevistado que puede responder “no sé”, “no me acuerdo”). El protocolo de entrevista de referencia del SVA es la *Entrevista Paso a Paso* (Step-wise interview) (Yuille Hunter, Joffe y Zaparniuk, 1993) en la que, tras requerir al testigo la narración en recuerdo libre, sigue un interrogatorio con preguntas generales (se pide al testigo que describa todo lo que recuerde sin obviar ningún detalle, se usan scripts); preguntas específicas (tienen como objeto aclarar información; si las preguntas implican una elección, ésta ha de ser de más de dos alternativas), y uso de ayudas a la entrevista (muñecos, dibujos u otros instrumentos, tras terminar todos los pasos anteriores, para clarificar alguna información). En la fase de interrogatorio es probable que se contamine la prueba, no resultando ni la concienciación, ni el entrenamiento de los entrevistadores suficientes para el control de esta fuente de contaminación (Pearse y Gudjonsson, 1996; Walsh y Milne, 2008). Como asume la propia técnica de entrevista, los interrogatorios están condicionados por scripts que implican una secuencia de los hechos que el entrevistador asume e impone en el relato. Pero la cuestión trasciende a la potencial contaminación de la prueba con el interrogatorio, al superar la tarea forense. La tarea forense no es de investigación (fact finder), que corresponde a policías, jueces y tribunales, sino de validación, en su caso, la memoria del testigo, como propia de hechos vividos (prueba indiciaria). No en vano, el procedimiento no valida un hecho o detalle concreto del relato, sino el conjunto de lo narrado. La técnica forense basada en el análisis de contenido de la declaración no tiene por objeto ni la investigación del caso, que corresponde a policías y operadores jurídicos, ni la exactitud, sino establecer la honestidad del testigo. De hecho, entre un 10 y un 20 % de la información honesta del testigo no es exacta (Köknken, Milne, Memon y Bull, 1999; Memon et al., 1997; Memon, Meissner y Fraser, 2010). La fase interrogativa es de investigación, aporta poca información (ceñida únicamente a la respuesta a la pregunta), no facilita la manifestación de criterios de contenido de memo-

rias de hechos vividos, al tiempo que es probable que contamine la prueba. Por tanto, debería obviarse en la técnica forense por falta de utilidad y riesgo de contaminación. No obstante, la entrevista también se puede contaminar con indicadores no verbales y paraverbales, que no implican un interrogatorio, por parte del entrevistador (e.g., aprobación, desaprobación, interrupción).

Un criterio jurisprudencial de evaluación de la credibilidad del testimonio es la persistencia en la incriminación (Sentencia 7070/1988 del TS, Sala de lo Penal, de 28 de septiembre de 1988). Este criterio supone que se han de tomar medidas repetidas del testimonio. El SVA no prevé la obtención de más de una declaración, por lo que toman las declaraciones obrantes en el procedimiento como medio de contraste de la obtenida como prueba forense. La persistencia se mide en el SVA a través de la (in)consistencia con las leyes naturales; (in)consistencia con otras declaraciones; e (in)consistencia con otras pruebas. Ahora bien, las declaraciones en condiciones de investigación no son comparables en contenido a las forenses (Köhnken et al., 1999; Memon et al., 1997, 2010), ni siquiera las pruebas preconstituidas, que no son más que interrogatorios judiciales hechos por peritos y que están afectados por los mismos condicionantes revisados.

Por su parte, los criterios jurisprudenciales y legales de revisión y réplica de la prueba ya fueron discutidos como criterios científicos.

Además de los criterios de realidad derivados de la hipótesis Undeutsch, los atributos de memoria procedentes del modelo del Reality Monitoring (RM) (Johnson y Raye, 1981), también se han estudiado para la discriminación entre memorias de origen externo (i.e., perceptuales, hechos vividos) y memorias de origen interno (esto es, imaginadas, fabricadas). Sucintamente, el modelo asume que los eventos percibidos contienen, por estar activada la memoria sensorial, más atributos sensoriales (e.g., información visual, auditiva), semánticos (e.g., tamaño, color) y contextuales (e.g., espacio, tiempo). Por el contrario, en las memorias externas predice una mayor presencia de operaciones cognitivas (e.g., pensamientos, razonamientos e inferencias hechas por otros). Los resultados de la investigación (ver en Tabla 6 el listado de

categorías; Sporer 1997, 2004) apoyan, en general, la efectividad de las categorías de contenido emanadas de este modelo (Masip, Sporer, Garrido, y Herrero, 2005; Vrij, 2008). No obstante, sólo algunos criterios discriminaron entre verdad y mentira; la puntuación total tampoco discernía sistemáticamente entre verdad y mentira; algunas categorías son escasamente útiles para esta tarea (e.g., información espacial); e incluso se han obtenido resultados contrarios a la predicción del modelo (e.g., operaciones cognitivas asociadas a memorias de origen externo) (Masip et al., 2005; Vrij, 2008).

Tabla 6

Criterios ampliados del RM de Sporer (1997, 2004)

Claridad (claridad, viveza en vez de vaguedad).
Información sensorial (tal como sonidos, gustos o detalles visuales).
Información espacial (lugares, ubicaciones).
Información temporal (ubicación del evento en el tiempo, descripción de secuencias de eventos).
Afecto (expresión de emociones y sentimientos sentidos durante el evento).
Reconstrucción de la historia (plausibilidad de reconstrucción del evento tras la información dada).
Realismo (plausibilidad, realismo y sentido de la historia).
Operaciones cognitivas (descripciones de inferencias hechas por otros durante el evento).

Construcción, propiedades psicométricas y criterios del Sistema de Evaluación Global para la clasificación de memorias de hechos vividos

Sobre la base de que ambos sistemas categoriales pueden complementarse (Sporer, 2004; Vrij, 2000, 2008) y que hay más categorías que las informadas en ambos modelos (Steller y Köhnken, 1989; Amado et al., 2016; Arce y Fariña, 2009), nos planteamos (e. g., Arce y Fariña, 2013a, 2014) construir un sistema categorial partiendo del CBCA, que contuviera categorías de realidad, atributos de memoria y otras surgidas de análisis de casos, que fuera metódico (i.e., fiable y válido) y del que resultara una técnica forense que cumpliera con los estándares Daubert, científicos y legales y jurisprudenciales

anteriormente revisados. De acuerdo con las directrices de Bardin (1996) para construir un sistema categorial fiable y válido (i.e., exclusión mutua, homogeneidad, objetividad, adecuación o pertinencia, exhaustividad, fidelidad y productividad), en primer lugar, construimos un manual con las definiciones de cada categoría, la unidad de análisis de cada categoría y ejemplos de cada una de ellas (objetividad). Las definiciones de las categorías,

dada la imprecisión y diferentes acepciones e interpretaciones a que daban lugar, fueron ajustadas con material de declaraciones procedentes de casos judiciales y fabricadas (objetividad). En concreto, se cotejaba, cuando había definiciones o interpretaciones diferentes, cuál (o la suma de ellas) era la que mejor discriminaba entre memorias (productividad). El conjunto de categorías resultante fue sometido a un análisis para el control de dupli-

Tabla 7

Categorías del sistema general (todo tipo de casos) de análisis de contenido de las declaraciones del SEG

CARACTERÍSTICAS GENERALES. Los criterios aquí englobados se refieren a la declaración tomada en su totalidad.
1.- Estructura lógica (incluye la reconstrucción de la historia y el realismo del RM, esto es, plausibilidad de reconstrucción del evento, realismo y sentido de la historia).
2.- Elaboración inestructurada.*
3.- Cantidad de detalles.*
4.- Concreción (SRA, concreción frente a estereotipos, incluye claridad del RM, esto es, viveza en vez de vaguedad).*
CONTENIDOS ESPECÍFICOS. Se evalúan partes específicas del testimonio referidas a la presencia o fuerza de ciertos tipos de descripciones.
5.- Engranaje contextual (incluye los criterios información espacial e información temporal del RM).
6.- Descripción de interacciones.
7.- Reproducción de conversaciones.*
8.- Complicaciones inesperadas durante el incidente
9.- Información perceptual (RM, esto es, información sensorial tal como sonidos, gustos o detalles visuales).
10.- Operaciones cognitivas (RM, codificación inversa).
PECULIARIDADES DEL CONTENIDO. Se incluyen aquí aquellas características de una declaración que aumentan su concreción o viveza.
11.- Detalles inusuales.*
12.- Detalles superfluos.*
13.- Incomprensión de detalles relatados con precisión.
14.- Originalidad de las expresiones (SRA).*
15.- Asociaciones externas relacionadas (se añade la información no relacionada directamente con los hechos, pero sí con el nodo de memoria recuperado).*
16.- Relatos del estado mental subjetivo (incluye el criterio afecto del RM).
17.- Atribución del estado mental del autor del delito.
CONTENIDOS REFERENTES A LA MOTIVACIÓN. Estos criterios desvelan la motivación del testigo para hacer la declaración.
18.- Correcciones espontáneas (se añade especificaciones y complementaciones de la información del SRA).*
19.- Admisión de falta de memoria.
20.- Plantear dudas sobre el propio testimonio.*
21.- Auto-desaprobación.
22.- Perdón al autor del delito.
ELEMENTOS ESPECÍFICOS DE LA AGRESIÓN. Elementos del testimonio que no se relacionan con la viveza general de la declaración, sino con el delito.
23.- Detalles característicos de la agresión (un ejemplo en casos de violencia de género sería que no es consciente de determinados tipos de agresiones, tal como las sexuales, pero describe relaciones sin consentimiento).
DAÑO PSICOLÓGICO DERIVADO DE LA ACCIÓN OBJETO DE JUICIO.
24.- Síntomas clínicos sutiles que forman parte de la huella psicológica (ver la relación de síntomas sutiles en Arce, Fariña, Carballal, y Novo, 2006, 2009).*

Nota. Las categorías del sistema han discriminado significativamente entre memorias de hechos vividos y fabricados (modelo productivo). **Criterios que discriminaron significativamente entre memorias de hechos vividos y fabricados con un tamaño del efecto moderado o grande (modelo robusto). RM = Criterios Ampliados del Reality Monitoring (Sporer, 1997); SRA = Statement Reality Analysis (Undeutsch, 1967, 1989). El modelo robusto reduce significativamente el número de falsos negativos en relación al modelo productivo. Se han encontrado categorías de análisis específicas de casuísticas concretas, tal como violencia de género, que se añaden a estas que no están afectadas por el contexto.

cidad de medidas (exclusión mutua) por medio de la técnica thurstone. La decisión acerca de definir las de forma excluyente o crear una categoría que englobara más de una categoría se tomó en función de estudio de la validez incrementada de clasificación con declaraciones de casos judiciales y fabricados. Se buscaron por medio de un procedimiento por aproximaciones sucesivas (ver procedimiento en Fariña et al., 2002) en memorias de casos reales (ground truth, sentencia judicial que clasificaba la declaración como verdadera no basada en una prueba de análisis de contenido de las declaraciones; memorias autobiográficas con imágenes de hechos no judicializados) nuevas categorías de contenido, hallándose categorías para todo tipo de casos y otras, específicas de casos (exhaustividad y pertinencia). Una vez identificadas, fueron sometidas al procedimiento anteriormente descrito, en el que se determinó si realmente era una nueva categoría y productiva, o formaba parte de una existente.

El sistema categorial resultante (ver Tabla 7) se mostró consistente internamente (homogeneidad), $\alpha = .789$ y $.856$ (Arce, Fariña, Novo, y Vilariño, 2013; Vilariño, Novo, y Seijo, 2011). Además, las categorías discriminaron (validez discriminante) significativamente entre memorias de hechos vividos y fabricados (Arce et al., 2013; Arce, Fariña, y Vivero, 2007; Vilariño, 2010; Vilariño et al., 2011). Se definió empíricamente (objetivo) un criterio de decisión estricto (validez criterial con falsos positivos = 0; principio de presunción de inocencia); se creó un sistema control de la correcta ejecución (fidelidad) de la medida en diseños de $N = 1$ (consistencia inter- e intra-evaluador, e inter-contextos) y se subsanaron los déficits del SVA. En concreto, la técnica forense, Sistema de Evaluación Global, exige que sea practicada por un perito titular entrenado en la misma; que se graben y guarden las declaraciones (revisión externa de los métodos usados y validez jurídica de la prueba); que se registren y guarden las codificaciones de las declaraciones (revisión externa de los registrados de los criterios y de la fiabilidad de éstos); que las codificaciones sean realizadas por dos peritos por separado (evaluación de la fiabilidad de la medida); que se recaben dos declaraciones (evaluación de la persistencia); y que las declaraciones se obtengan por medio de la entrevista cognitiva –control de contaminación de la prueba con el interrogatorio–: recuerdo libre

y técnicas de ayuda de recuperación de memoria (sin interrogatorios). La evaluación de la persistencia, validez de la prueba, se obtiene si cumple los siguientes cinco criterios tomados de análisis de sentencias judiciales (Arce, Seijo y Novo, 2010): consistencia interna (¿tiene contradicciones internas en el relato?); consistencia externa (¿es consistente con otras pruebas robustas o incontrovertibles?); consistente con la anterior(es) (¿hay consistencia en la información central inter-declaraciones?); persistencia (¿son estables las declaraciones en el tiempo en los contextos?); y consistencia con las leyes científicas y de la naturaleza (¿contiene el relato hechos incompatibles con las leyes científicas o de la naturaleza?). Si la prueba incumple alguno de estos criterios de validez de la misma, no se realiza un análisis de contenido por ser la prueba inválida. Además, se le requiere que tenga la suficiente amplitud para ser objeto del análisis de contenido, prueba suficiente (¿Supera la capacidad de memoria del testigo? ¿Contiene toda la información necesaria del evento?).

Líneas futuras de investigación

Sobre la base de que las categorías de contenido discriben entre las memorias de hechos vividos o auto-experimentados y fabricados; que se pueden construir sistemas categoriales metódicos; que puede haber más categorías, especialmente propias de contextos de victimización específicos i.e., acoso escolar, violencia de género, acoso laboral, agresiones sexuales a adultos (exhaustividad), la investigación ha de dirigirse a la detección y estudio de nuevas categorías de contenido en esos contextos. El método más eficaz y productivo, definido previamente, es el de aproximaciones sucesivas tomando como material declaraciones procedentes de casos judiciales y fabricadas, y aplicando a las categorías identificadas las técnicas descritas para cumplir con las directrices de Bardin para crear un sistema categorial metódico. Asimismo, el criterio de decisión estricto (que garantice que los falsos positivos sean 0) ha de ser revisado continuamente pues la mentira se aprende (importancia de no enseñar las claves de la técnica para minimizar el aprendizaje de la misma), en tanto la honestidad se mantiene constante.

Referencias

- Amado, B. G., Arce, R. y Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A Meta-Analytic Review. *European Journal of Psychology Applied to Legal Context*, 7, 3-12. <http://dx.doi.org/10.1016/j.ejpal.2014.11.002>
- Amado, B. G., Arce, R., Fariña, F. y Vilariño, M. (2016). CBCA reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16, 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- Arce, R. (2015). *Manifesto de rexeitamento da violencia de xénero. 25 de novembro 2015 [Manifesto of rejection of gender violence. November 25, 2015]*. Recuperado de https://xornal.usc.es/xornal/opinions/opinion_0175.html
- Arce, R. (2017, mayo). Criterios Daubert, científicos, jurisprudenciales, procesales y legales a aplicar a la prueba psicológica forense [Daubert Criteria, Scientific, Jurisprudential, Procedural, and Legal to Apply to the Forensic Psychological Test]. En *Taller en el X Congreso Internacional de Psicología Jurídica y Forense*, Sevilla, España.
- Arce, R. y Fariña, F. (2013a). Evaluación psicológica-forense en casos de violencia de género: El Sistema de Evaluación Global (SEG). En S. P. Colín, E. García-López, y L. A. Morales (Coords.), *Ecos de la violencia, voces de la reconstrucción* (pp. 95-120). Morelia, Michoacán, México: Servicio de Publicaciones de la Universidad de Morelia. <https://doi.org/10.13140/RG.2.1.1885.1040>.
- Arce, R. y Fariña, F. (2013b). Psicología forense experimental. Testigos y testimonio. Evaluación cognitiva de la veracidad de testimonios y declaraciones [Experimental Forensic Psychology. Witnesses and Testimony. Cognitive Evaluation of the Accuracy of Testimonies and Statements]. En S. Delgado (Dir. Tratado), y S. Delgado, y J. M. Maza (Coords., Vol. V), *Tratado de Medicina Legal y Ciencias Forenses: Vol. V. Psiquiatría legal y forense* (pp. 21-46). Barcelona, España: Bosch.
- Arce, R. y Fariña, F. (2014). Práctica de la prueba psicológica-forense: El Sistema de Evaluación Global (SEG) [Practice of Forensic Psychological testing: The Global Assessment System (GAS)]. En C. Chan, C. Estrada, y F. J. Rodríguez (Eds.), *aportaciones a la Psicología Jurídica y Forense desde Iberoamérica* (pp. 47-61). México, DF: Manual Moderno.
- Arce, R., Fariña, F., Carballal, A. y Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación [Evaluating Psychological Injury in Motor Vehicle Accidents (MVA): Development and Validation of a Protocol for Detecting Simulation]. *Psicothema*, 18, 278-283. Recuperado de <http://www.psicothema.com/pdf/3210.pdf>
- Arce, R., Fariña, F., Carballal, A. y Novo, M. (2009). Creación y validación de un protocolo de evaluación forense de las secuelas psicológicas de la violencia de género [Creation and Validation of a Forensic Protocol to Assess Psychological Harm in Battered Women]. *Psicothema*, 21, 241-247. Recuperado de <http://www.psicothema.es/pdf/3621.pdf>
- Arce, R., Fariña, F., Novo, M. y Vilariño, M. (2013). Contraste de la efectividad en la detección de declaraciones fabricadas de los criterios de validez del Sistema de Evaluación Global en casos de violencia contra la mujer [Contrast of effectiveness in detecting manufactured statements against the Global Assessment System's validity criteria in cases of violence against women]. En F. Expósito, I. Valor-Segura, M. Vilariño y A. Palmer (Eds.), *Psicología Jurídica aplicada a los problemas sociales* (pp. 153-161). Santiago de Compostela, España: Sociedad Española de Psicología Jurídica y Forense.
- Arce, R., Fariña, F. y Vivero, A. (2007). Estudio exploratorio de la efectividad de las técnicas de

- análisis de contenido de las declaraciones en casos de violencia de género [Exploratory Study of the Effectiveness of the Techniques of Content Analysis of Declarations in Cases of Gender-Based Violence]. En C. Guillén y R. Guil (Coords.), *Psicología social: Un encuentro de perspectivas* (Vol. 1, pp. 590-604). Cádiz, España: Asociación de Profesionales de la Psicología Social.
- Arce, R., Seijo, A. y Novo, M. (2010). Testimony Validity: A Comparative Study of Legal and Empirical Criteria. *Psychology in Spain*, 14, 1-7. Recuperado de <http://www.psychologyinspain.com/content/full/2010/14010.pdf>
- Arntzen, F. (1970). *Psychologier der Zeugenaussage* [Psychology of Testimony]. Göttinger, Alemania: Hogrefe.
- Bardin, L. (1996). *El análisis de contenido* [Content analysis] (2a. ed.). Madrid, España: Akal.
- Boente, S. E. R. (2011). La prueba en los supuestos de violencia de género [Evidence in GBV Assumptions]. *Revista Iberoamericana de Estudios Utilitaristas*, XVIII(1-2), 231-246. Recuperado de <http://www.usc.es/revistas/index.php/telos/article/download/1413/1656>
- Campbell, D. T. (1969). Reforms as Experiments. *American Psychologist*, 24(4), 409. <https://doi.org/10.1037/h0027982>
- Daubert vs. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786 (1993).
- Dettenborn, H. Froehlich, H. y Szewczyk, H. (1984). *Forensische Psychologie* [Forensic Psychology]. Berlín, Alemania: Deutscher Verlag der Wissenschaften.
- Fariña, F., Arce, R. y Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in Judicial Decision Making]. *Psicothema*, 14(1), 39-46. Recuperado de <http://www.psicothema.com/pdf/684.pdf>
- Fariña, F., Arce, R. y Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Linepus: A Comparison of High-Fidelity Research and Research in a Real Context]. *Psicothema*, 6, 395-402. Recuperado de <http://www.psicothema.es/pdf/935.pdf>
- Hans, V. P. y Vidmar, N. (1986). *Judging the Jury*. New York, NY: Plenum Press.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O. y Reiter-Lavery, L. (1997). Reliability of Criteria-Based Content Analysis of Child Witness Statements. *Legal and Criminological Psychology*, 2, 11-21. <https://doi.org/10.1111/j.2044-8333.1997.tb00329.x>
- Johnson, M. K. y Raye, C. L. (1981). Reality Monitoring. *Psychological Review*, 88, 67-85. <https://doi.org/10.1037/0033-295X.88.1.67>
- 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, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511490071.003>
- Köhnken, G., Milne, R., Memon, A. y Bull, R. (1999). The Cognitive Interview: A Meta-analysis. *Psychology, Crime & Law*, 5, 3-27. <https://doi.org/10.1080/10683169908414991>
- Konecni, V. J. y Ebbesen, E. B. (1992). Methodological Issues on Legal Decision-Making, with Special Reference to Experimental Simulations. En F. Lösel, D. Bender y T. Bliesener (Eds.), *Psychology and Law: International Perspectives* (pp. 413-423). Berlín, Alemania: Walter de Gruyter.
- Masip, J., Sporer, S. L., Garrido, E. y Herrero, C. (2005). The Detection of Deception with the Reality Monitoring Approach: A Review of the Empirical

- Evidence. *Psychology, Crime & Law*, 11, 99-122. <https://doi.org/10.1080/10683160410001726356>
- Memon, A., Meissner, C. A. y Fraser, J. (2010). Cognitive interview. A Meta-Analytic Review and Study Space Analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16, 340-372. <https://doi.org/10.1037/a0020518>
- Memon, A., Wark, L., Bull, R. y Köhnken, G. (1997). Isolating the effects of the Cognitive Interview Techniques. *British Journal of Psychology*, 88(2), 179-198. <https://doi.org/10.1111/j.2044-8295.1997.tb02629.x>
- Ministerio de Justicia. (2005). *Guía y manual de valoración integral forense de la violencia de género y doméstica*. Madrid, España [Guide and Manual for the Comprehensive Forensic Assessment of Domestic and Gender-Based Violence]: Secretaría General Técnica del Ministerio de Justicia. Recuperado de [http:// goo.gl/7dVRxN](http://goo.gl/7dVRxN)
- Novo, M. y Seijo, D. (2010). Judicial Judgement-Making and Legal Criteria of Testimonial Credibility. *European Journal of Psychology Applied to Legal Context*, 2, 91-115. Recuperado de http://sepjf.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=26&Itemid=110&lang=en
- Pearse, J. y Gudjonsson, G. H. (1996). Police Interviewing Techniques at Two South London Police Stations. *Psychology, Crime and Law*, 3, 63-74. <https://doi.org/10.1080/10683169608409795>
- Popper, K. R. (1961). *The Logic of Scientific Discovery* (2a. ed.). New York, NY: Harper and Row.
- Roma, P., San Martini, P., Sabatello, U., Tatarelli, R. y Ferracuti, S. (2011). Validity of Criteria-Based Content Analysis (CBCA) at trial in free-narrative interviews. *Child Abuse & Neglect*, 35, 613-620. <https://doi.org/10.1016/j.chiabu.2011.04.004>
- Sentencia 210/2014 del TS, Sala de lo Penal, de 14 de marzo de 2014. Recuperado de <https://supremo.vlex.es/vid/abuso-sexual-victima-declaraciones-503438218>
- Sentencia 213/2002 del TS, Sala de lo Penal, 14 de febrero de 2002. Recuperado de <http://supremo.vlex.es/vid/agresion-sexual-absolucion-fa-u-15055496>
- Sentencia 246/2016 de la AP de Las Palmas de Gran Canaria, Sala de lo Penal, de 28 de junio de 2016. Recuperado de <https://www.iberley.es/jurisprudencia/sentencia-penal-n-246-2016-ap-palmas-sec-1-rec-62-2015-28-06-2016-47669530>
- Sentencia 7070/1988 del TS, Sala de lo Penal, de 28 de septiembre de 1988. Recuperado de <http://lawcenter.es/w/file/download/64666>
- 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. [https://doi.org/10.1002/\(SICI\)1099-0720\(199710\)11:5<373::AID-ACP461>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0)
- Sporer, S. L. (2004). Reality monitoring and detection of deception. En A. Granhag y L. A. Strömwall (Eds.), *The Detection of Deception in Forensic Contexts* (pp. 64-102). Cambridge, UK: Cambridge University Press.
- Steller, M. (1989). Recent Developments in Statement Analysis. En J. C. Yuille (Ed.), *Credibility Assessment* (pp. 135-154). Dordrecht, Holanda: Kluwer Academic Publishers.
- Steller, M. y Böhm, C. (2006). 50 años de jurisprudencia del Tribunal Supremo alemán sobre psicología del testimonio. Balance y perspectiva [50 Years of German Supreme Court Case Law on Witness Psychology. Balance and Perspective]. En T. Fabian, C. Böhm y J. Romero (Eds.), *Nuevos*

- caminos y conceptos en la psicología jurídica* (pp. 53-77). Berlín, Alemania: Lit Verlag.
- Steller, M. y Köhnken, G. (1989). Criteria-Based Content Analysis. En D. C. Raskin (Ed.), *Psychological methods in Criminal Investigation and Evidence* (pp. 217-245). New York, NY: Springer-Verlag.
- Sooniste, T., Granhag, P. A., Strömwall, L. A. y Vrij, A. (2015). Statements about True and False Intentions: Using the Cognitive Interview to Magnify the Differences. *Scandinavian Journal of Psychology*, 56(4), 371-378. <https://doi.org/10.1111/sjop.12216>
- Steller, M. y Köhnken, G. (1989). Criteria-Based Content Analysis. En D. C. Raskin (Ed.), *Psychological Methods in Criminal Investigation and Evidence* (pp. 217-245). New York, NY: Springer-Verlag.
- Szewczyk, H. (1973). *Kriterien der beurteilung kindlicher zeugenaussagen* [Criteria for Assessing the Validity of Child Declarations]. *Probleme und Ergebnisse der Psychologie*, 46, 46-66.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen [Assessing the Credibility of Witnesses]. En U. Undeutsch (Ed.), *Handbuch der psychologie*, Vol. II: Forensische psychologie (pp. 26-181). Göttingen, Alemania: Verlag für Psychologie.
- Undeutsch, U. (1989). The Development of Statement Reality Analysis. En J. Yuille (Ed.), *Credibility Assessment* (pp. 101-119). Dordrech, Holanda: Kluwer Academic Publishers.
- Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* [Is it Possible to Discriminate Real Statements from Imagined and Real Psychological Footprint of Simulated in Cases of Gender-Based Violence?] (Tesis doctoral, Universidade de Santiago de Compostela, España). Recuperado de <http://goo.gl/rPD728>
- Vilariño, M., Novo, M. y Seijo, D. (2011). Estudio de la eficacia de las categorías de realidad del testimonio del Sistema de Evaluación Global (SEG) en casos de violencia de género [Study of the Effectiveness of the Reality Categories of Global Evaluation System (GAS) Testimony in Cases of Gender-Based Violence]. *Revista Iberoamericana de Psicología y Salud*, 2, 1-26. Recuperado de <http://www.redalyc.org/pdf/2451/245116403001.pdf>
- Volbert, R. y Steller, M. (2014). Is this Testimony Truthful, Fabricated, or Based on False Memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19, 207-220. <https://doi.org/10.1027/1016-9040/a000200>
- Vrij, A. (2000). *Detecting Lies and Deceit*. Chichester, UK: John Wiley and Sons.
- Vrij, A. (2005). Criteria-based Content Analysis. A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, and Law*, 11, 3-41. <https://doi.org/10.1037/1076-8971.11.1.3>
- Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2a. ed.). Chichester, UK: John Wiley and Sons.
- Walsh, D. W. y Milne, R. (2008) Keeping the PEACE? An Analysis of the Taped Interview Performance of Benefit Fraud Investigators Within the DWP. *Legal and Criminological Psychology*, 13, 39-57. <https://doi.org/10.1348/135532506X157179>
- Weick, K. E. (1985). Systematic Observational Methods. En G. Lindzey y E. Aronson (Eds.), *The Handbook of Social Psychology* (Vol. 1, pp. 567-634). Hillsdale, NJ: LEA.
- Yuille, J.C., Hunter, R., Joffe, R. y Zaparniuk, J. (1993). Interviewing Children in Sexual Abuse Cases. En G. Goodman y B. Bottoms (Eds.), *Child Victims, Child Witnesses: Understanding and Improving Children's Testimony* (pp. 95-115). Nueva York, NY: Guilford Press.

CONTENT ANALYSIS OF THE WITNESS STATEMENTS: EVALUATION OF THE SCIENTIFIC AND JUDICIAL VALIDITY OF THE HYPOTHESIS AND THE FORENSIC PROOF

RAMÓN ARCE

EXTENDED SUMMARY

Introduction

The evaluation of the veracity of a testimony is the cornerstone on which most judicial decisions rest on (Novo & Seijo, 2010), particularly in cases of crimes committed in the private and domestic sphere such as sexual and violent crimes (e.g., family and gender violence), which are characterised by contradictory versions of events. This contingency has been estimated to affect approximately 85 % of trials worldwide (Hans & Vidmar, 1986). In order for one version to prevail over the other or to confer credibility to a victim's-plaintiff's testimony where circumstances may concur to undermine the credibility of a testimony (e.g., resentment, hostility, financial gain, conflicts of interest), it must be ratified and substantiated by admissible evidence.

Evaluation of the Undeutsch hypothesis under the Daubert standard

In the case of *Daubert v. Merrell Dow Pharmaceuticals* (1993), the US Supreme Court established the criteria for the admission of expert testimony as scientific evidence in a court of law, which has also been tested in scientific research: 1) Is the scientific hypothesis/technique testable? Not only is it possible to test it, it has already been tested in both experimental (laboratory) studies, and field studies. As for the validation of a theory (hypothesis), the US Supreme Court ruled that validat-

ing a theory only on the findings of experimental studies was insufficient, and it should be further substantiated by field studies. Both of the study designs have advantages and shortcomings. The advantage of experimental studies is the high degree of control over the independent variable (causal) manipulated by the researcher, meaning the reliability of the classification of real or fabricated accounts is total at the cost of external validity, particularly in the field of research on the psychology of testimonies. Thus, a study comparing testimonies in a high fidelity experimental study and a field study found inconsistent results (Fariña, Real, & Arce, 1992). The strength of field studies lies in the external validity of the designs, but their weakness is internal validity as they lack control over the manipulated variables i.e., the classification criterion (ground truth) of real events. 2) Has the hypothesis/technique been tested? It has been widely tested and validated by numerous studies (i.e., it is statistically significant in discriminating between memories of real-life events and fabricated accounts), both in experimental studies ($\delta = 0.56$ and 0.32 , in populations of children and adults, respectively), and field studies ($\delta = 2.71$ and 0.45 ; in children and adults, respectively); in plaintiff testimonies in populations of children ($\delta = 0.79$), and adults ($\delta = 0.56$); in eyewitnesses ($\delta = 0.51$) and defendants ($\delta = 0.35$); in an array of crimes (e.g., child sex abuse, $\delta = 0.79$; sexual assault and gender violence, $\delta = 0.87$); and in several contexts (memories of criminal and non-criminal events; memories of fabricated events, induced or suggested accounts) (Amado et al., 2015, 2016). The validity of the hypothesis in both field and experimental studies underscores its robust study design (convergent validity). In other words, it is robust to sources of error

linked to the design of experimental studies (external validity), and field studies (internal validity). Moreover, it is equally valid in different contexts (i.e., types of crimes), and exhibits inter-context consistency. In addition, the hypothesis has not been rejected by inconvenient results i.e., negative or adverse to the hypothesis mean effect sizes. Likewise, the ground truth factor does not mediate effects in results acquiring consistency among ground truth standards. Thus, the hypothesis has not only been tested and substantiated, it has also been validated and is powerful (moderate, large and more than large effect sizes in different settings). 3) Is there a known error rate? Given that the hypothesis was formulated but not corroborated by the author, that the categorical systems were created on the basis of the authors' experiences (top-down) derived from thousands of cases without a strict and objective decision criterion, and that it was not tested, would explain why the error rate was not published. Nevertheless, the meta-analysis of Amado et al. (2015, 2016) calculated the statistical margin of error following a systematic analysis of the literature to be around 31.5% of false negatives in the population of children and 36.5% in adults. 4) Has the hypothesis/technique been subjected to peer review and publication (it is not a prerequisite 'sine qua non')? On more than 800 occasions (effect size) (Amado et al., 2015, 2016). 5) Is the hypothesis upon which the technique rests on generally accepted as appropriate by the scientific community? The US Supreme Court resorts to the scientific community to attest the method and principles of the hypothesis. To our knowledge, there has been no surveys regarding this or any other psychological hypothesis. From a scientific point of view, the means for consulting the scientific community involves a systematic, meta-analytical review of the literature to determine broad support for the hypothesis i.e., validation by the scientific community (general acceptance). The meta-analysis of Amado et al. (2015, 2016) not only reveals general acceptance, but also found no meta-analytical result (i.e., mean effect size) contradicting the hypothesis.

Evaluation of the content analysis technique of SVA/CBCA statements in scientific, legal, and jurisprudential criteria

Compliance of the hypothesis with the Daubert standard does not intrinsically imply that the same is applicable to the technique derived from it (ecological fallacy), nor do the criteria stipulated by the US Supreme Court satisfy the mandatory requirements for forensic evidence (technique); thus, the need for additional scientific (see Table 1), legal, and jurisprudential criteria as shown in Table 2 (Arce, 2017). Of the forensic techniques derived from the Undeutsch hypothesis, SVA/CBCA (Steller & Köhnken, 1989), is the reference of choice, and is a product of combining previous techniques (Arntzen, 1970; Dettenborn, Froehlich, and Szweczyk, 1984; Szweczyk, 1973; Undeutsch, 1967). SVA describes the technical procedure (see Table 3) and provides a validity checklist (see Table 5), whereas CBCA outlines the reality criteria (see Table 4). As for the scientific criteria applied to the forensic technique, the categorical system of reality criteria, CBCA is not methodic i.e., it is neither reliable nor valid as it has no means for estimating the correct application of the technique to a specific case; and there is no strict and objective decision criterion, making the decision semi-objective or semi-standardized. In relation to legal and jurisprudential criteria, the technique does not safeguard the constitutional principle of the presumption of innocence (translated into the jurisprudential maxim that a person is considered innocent unless proven guilty, that is, guilt must be proved beyond reasonable doubt, or otherwise the accused must be acquitted); the technique does not meet legal and jurisprudential standards for taking statement given that the semi-structured interview includes an interrogation that can contaminate the evidence; without evaluating the consistency of the testimony (only one statement is obtained and analysed). In short, the underlying Undeutsch hypothesis is valid, but the technique fails to meet the scientific, legal, and jurisprudential criteria.

Design, psychometric properties, and criteria of the Global Evaluation System for the classification of memories of real-life events

Besides the Undeutsch hypothesis, the Reality Monitoring model (Johnson & Raye, 1981), has also proven to be valid for discriminating between external (i.e., perceived, real-life events) and internal memories (i.e., fabricated or fictitious accounts). Succinctly, certain categories of content analysis significantly discriminated between both types of memory, whilst others did not or were unproductive, and some were even contrary to the model (Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2008). By combining categories from both models and adding more content categories of statements, a reliable and valid methodic categorial system was designed based on CBCA, the combination of reality categories, memory attributes and others derived from the analysis of cases to ensure the forensic technique complies with the Daubert standards, as well as with the scientific, legal, and jurisprudential requirements of a forensic technique (Arce and Fariña, 2013, 2014).

The design of the categorial system was in accordance with the guidelines of Bardin (1996) to ensure it was methodic i.e., reliable and valid: mutual exclusion, homogeneity, objective, adequate and relevant, exhaustive, fidelity and productivity. The resulting categorial system (see Table 7) was internally consistent (homogeneity), $\alpha = .789$ and $.856$ (Arce, Fariña, Novo, & Vilariño, 2013; Vilariño, Novo, & Seijo, 2011). Moreover, the categories discriminated (discriminant validity) significantly between memories of real-life events and fabricated accounts (Arce et al., 2013; Arce, Fariña, & Vivero, 2007; Vilariño, 2010; Vilariño et al., 2011). A strict decision criterion was empirically (objectively) defined (criterion validity with false positives = 0; principle of the presumption of innocence); a correct execution control system was designed (fidelity) of the measure in the designs of $N = 1$ (inter- and intra-rater, and inter-context consistency) and to overcome the deficiencies of SVA. In fact, the Global Evaluation System is a forensic technique that must be carried out by a trained expert; statements must be recorded and archived (external review of

the methods used and judicial validity of the evidence); the coding of statements must be registered and archived (external review of the registered criteria and their reliability); codings must be carried out independently by two experts (evaluation of the reliability of the measure); two statements are obtained (evaluation of the consistency of the evidence); and that statements are obtained through a cognitive interview –control of the contamination of evidence resulting from the interrogation–: free-recall and techniques to aid memory recall (without interrogation). The evaluation of the validity of the testimony (validity of the evidence) is ensured if the following 5 criteria based on the analysis of judicial judgements are met (Arce, Seijo, & Novo, 2010): internal consistency (Are there internal contradictions in the account?); external consistency (Is it consistent with robust or other irrefutable evidence?); consistency with prior and subsequent testimonies (Is there consistency between the central aspects and actions of the event?); persistence (Are statements stable in time in the contexts?); and consistent with scientific laws and the laws of nature (Does the account mention events incompatible with scientific laws and the laws of nature?). If the evidence fails to meet any of these validity criteria, no content analysis is undertaken as the evidence is invalid. Moreover, there should be sufficient evidence for it to be the object of content analysis, sufficient evidence (Is it beyond the witness' memory capacity? Does it contain all the necessary information regarding the events?).

Further lines of research

Taking into account content categories discriminate between memories of self-experienced real-life events versus fabricated or fictitious accounts; that a methodic categorial systems may be designed; that there can be more categories, mainly those specific to contexts of victimization i.e., school bullying, gender violence, harassment at work, sexual assault in adults (exhaustive), research should seek to detect and analyse new content categories for the previously mentioned contexts. The most efficacious and productive method entails successive approximations (Fariña, Arce, & Novo, 2002) based on statements from judicial cases and fabricated

accounts, and applying the techniques described to the categories identified in accordance with the guidelines of Bardin for designing a methodic categorial system. Furthermore, a strict decision criterion (to ensure the number of false positives is 0) must be continuously revised since lying is also subject to continuous evolution (thus the importance of not revealing intricacies of the technique to prevent the learning of the technique); in comparison, honesty always remains constant.