



¿Son los criterios de realidad válidos para clasificar y discernir entre memorias de hechos auto-experimentados y de eventos vistos en vídeo?

Andrea Monteiro¹, María José Vázquez², Dolores Seijo^{3*} y Ramón Arce³

¹Unidad de Psicología Forense, Universidad de Santiago de Compostela, ²Departamento Aipse, Universidad de Vigo y

³Departamento de Ciencia Política y Sociología

• Recibido: 03 - 05 - 2018 • Aceptado: 28 - 05 - 2018 • Avance online: 04 - 06 - 2018

RESUMEN: El análisis de contenido de las declaraciones basado en criterios de realidad (hipótesis Undeutsch) se ha mostrado como una herramienta válida para la discriminación entre memorias de hechos auto-experimentados y memorias fabricadas. Sin embargo, numerosa evidencia científica proviene de experimentos en los que las memorias reales fueron sustituidas por memorias de hechos visionados. No existe evidencia para sostener que los criterios de realidad discriminen entre memorias de hechos auto-experimentados y fabricadas a partir de un evento visto en vídeo. Para conocer si los criterios de realidad discernen entre ambas memorias, se obtuvo el testimonio (entrevista cognitiva) de 30 adultos de una memoria autobiográfica (auto-experimentado) y una memoria fabricada a partir del visionado de un vídeo. Los resultados mostraron que el total de criterios de realidad del CBCA (Criteria-Based Content Analysis), así como ciertos criterios de realidad discriminaron entre ambas memorias en línea con la hipótesis Undeutsch. Otros criterios no discriminaron, resultaron improductivos e incluso contrarios a la hipótesis. Un patrón similar fue encontrado para la tasa de la memoria auto-experimentada. Se discuten las implicaciones de los resultados para la práctica forense.

PALABRAS CLAVE: Criteria-Based Content Analysis (CBCA), Credibilidad, Testimonio, Memoria de hechos auto-experimentados, Memoria fabricada de hechos, Criterios de realidad.

Are the reality criteria valid to classify and to discriminate between memories of self-experienced events and memories of video-observed events?

ABSTRACT: Statement content analysis based on reality criteria (Undeutsch hypothesis) has revealed as a valid instrument to discriminate between memories of self-experienced events (real-life) and fabricated memories of accounts. Nevertheless, much of the available scientific evidence provides from experiments where real-life memories were substituted by memories of events seen video (non-experienced). As for knowing if reality criteria do discriminate between self-experienced memories and fabricated memories of events seen in a video, the testimony of 30 adult participants was obtained about an autobiographic event (self-experienced) and about of an event seen on video (fabricated). The results showed that the total CBCA (Criteria-Based Content Analysis) criteria, as well as several reality criteria discriminated between both memories in line with the Undeutsch hypothesis. Nonetheless, other reality criteria did not discriminate, were unproductive or discriminated against the hypothesis. A similar pattern of outcomes was found for the classification rate of the self-experienced memory. The implications of the results for forensic practice are discussed.

KEYWORDS: Criteria-Based Content Analysis (CBCA), Credibility, Testimony, Memory of self-experienced events, Fabricated memories of events, Reality criteria.

La Teoría del Procesamiento de la Información ha sido adaptada con éxito a la formación de juicios en el contexto judicial (Kaplan y Anderson, 1973). En dicha adaptación, las pruebas son evaluadas en dos dimensiones: la fiabilidad y

*Correspondence: Dolores Seijo.

Departamento de Ciencia Política y Sociología. Universidad de Santiago de Compostela
Código Postal: 15782, Santiago de Compostela, España.
E-mail: mariadolores.seijo@usc.es

la validez (Ostrom, Werner, y Saks, 1978). La fiabilidad está demarcada por factores como la credibilidad del testimonio, la consistencia de las pruebas o la probabilidad de ocurrencia del evento, en tanto la validez por la relevancia de la prueba para la formación del juicio. De este modo, la credibilidad del testimonio se ha encontrado que es la piedra angular para la formación de juicios legales, especialmente en delitos cometidos en la esfera privada como los delitos sexuales o en el ámbito familiar (Arce,

Fariña, y Fraga, 2000; Boccaccini, 2002) en los que el testimonio de la víctima-denunciante (fiabilidad) es la prueba central (frecuentemente el testimonio de la víctima es la única prueba para sustentar la acusación) de la acusación (validez). Una vez que el testimonio de la víctima-denunciante es categorizado como fiable, éste sirve como anclaje que dirige el procesamiento de la información, y, subsecuentemente, la formación de juicios (Fariña, Arce, y Novo, 2002). Para ello, la jurisprudencia ha establecido que el testimonio de la víctima-denunciante para desvirtuar por sí mismo la presunción de inocencia ha de satisfacer los siguientes tres requisitos: ausencia de incredibilidad subjetiva (inexistencia de un móvil o interés que reste credibilidad a la declaración), verosimilitud (pruebas periféricas que corroboren el testimonio de la víctima) y persistencia en la incriminación (declaración estable en el tiempo y sin ambigüedades). El criterio ausencia de incredibilidad subjetiva suele debilitarse por la concurrencia de alguna motivación para la denuncia de modo que el testimonio de la víctima-denunciante pierde la objetividad requerida para convertirse en testimonio de parte. Dicha debilidad no invalida, per se, la declaración, pudiendo compensarse con una fortaleza en los otros criterios. Ésta proviene de pruebas periféricas que doten de valor al testimonio de la víctima-denunciante. El daño, fundamentalmente el físico y psicológico, y la evaluación de la fiabilidad del testimonio son las pruebas que, acorde al estudio sistemático de las sentencias judiciales, confieren valor de prueba suficiente al testimonio de la víctima-denunciante para desvirtuar la presunción de inocencia, dado que la falta de consistencia interna y externa, así como en los elementos centrales del testimonio de la víctima-denunciante conllevan de forma sistemática a la absolución del encausado (Novo y Seijo, 2010). La prueba admitida judicialmente para dotar de verosimilitud al testimonio de la víctima-denunciante en muy diversas legislaciones (e.g., España, Holanda, EE.UU, Alemania, Suecia) es la basada en el análisis de contenido del testimonio, principalmente la técnica Statement Reality Analysis [SVA] (Arce, 2017; Steller y Böhm, 2006; Vrij, 2008). Esta se sustenta en la conocida como hipótesis Undeutsch que establece que la memoria de hechos vividos (auto-experimentados), esto es, de un evento real difiere en contenido y calidad de una memoria no experimentada de

un evento (fabricada) (Undeutsch, 1967, p. 125). A partir de esta hipótesis, Undeutsch creó un sistema categorial de análisis de contenido, el Statement Reality Analysis [SRA], en el que identificaba aquellas categorías de contenidos de la declaración propias, que no exclusivas, con el que poder clasificar las declaraciones basadas en hechos vividos. Posteriormente, Steller y Köhnken (1989) sistematizaron y ampliaron ese sistema categorial en el Criteria Based Content Analysis [CBCA] que se enmarca en la técnica del SVA. Si bien, el sistema categorial CBCA no es metódico, dando lugar a una prueba que no sólo no cumple con los criterios científicos exigibles (fiabilidad, validez, criterio de decisión objetivo y estricto y carencia de control de la fidelidad de la aplicación de la medida), sino también procesales y jurisprudenciales (no evalúa la persistencia en el testimonio; la aplicación de la técnica contamina la prueba; no controla los falsos positivos —vulneración del principio de presunción de inocencia—) (Arce, 2017). Meta-análisis recientes han encontrado que las categorías de realidad de dicho sistema discriminan significativamente (y, por tanto, clasifican por separado, aunque no todas, y en conjunto), entre memorias de eventos vividos (auto-experimentados) y memorias fabricadas de eventos (no experimentados) en diferentes poblaciones (niños/as y adultos/as), contextos (e.g., abusos sexuales, violencia doméstica) y condiciones de testimonio (e.g., víctima-denunciante, denunciado, testigos) (Amado, Arce, y Fariña, 2015; Amado, Arce, Fariña, y Vilariño, 2016; Oberlader et al., 2016). En suma, la hipótesis es válida, pero no así el sistema categorial como conjunto. Además, permanece invariable en las condiciones experimentales mencionadas (generalizable), que incluye no sólo hechos auto-experimentados como propios (testigo participante, tal como testigo-víctima y denunciado) sino también como testigo (testigo presencial). No obstante, no hay suficiente evidencia científica para generalizar la validez de la hipótesis a la discriminación entre memorias de hechos auto-experimentados y no experimentados, pero visionados en vídeo. Sólo un estudio de Lee, Klaver y Hart (2008) abordó este diseño ($N = 45$), hallando que las categorías de realidad del CBCA engranaje contextual ($r = .40$), reproducción de conversaciones ($r = .30$), complicaciones inesperadas ($r = .51$), y atribución al estado mental del autor del delito

($r = .30$) fueron codificadas significativamente más en declaraciones de hechos vividos, confirmando la hipótesis, en tanto la categoría correcciones espontáneas ($r = .39$), contrariamente a la hipótesis, se registró más memorias de hechos no experimentados y visionados en vídeo que en memorias de hechos auto-experimentados. Por el contrario, mucha de la literatura que se aporta en apoyo de la capacidad de discriminación de los criterios de realidad del CBCA utiliza diseños que comparan memorias de hechos no experimentados visionados en vídeo, como si fueran auto-experimentados, con memorias fabricadas de hechos (Amado et al., 2016). En todo caso, estas mismas revisiones meta-analíticas encontraron que los criterios de realidad igualmente discriminaban entre memorias de hechos realmente vividos (auto-experimentados) y memorias fabricadas (estudios de campo).

A fin de aportar más evidencia sobre la validez de las categorías de realidad para discernir entre memorias de eventos vividos y memorias de eventos visionados, nos planteamos un estudio con el que contrastar la eficacia discriminativa de las categorías de realidad del CBCA entre memorias de hechos vividos (autobiográficas) y memorias fabricadas de hechos basadas en el visionado de un vídeo.

MÉTODO

• PARTICIPANTES

La muestra utilizada en el presente estudio se compuso de 30 adultos, 18 mujeres (67%) y 12 hombres (33%), $\chi^2(1) = 1.20$, ns, con edades comprendidas entre 18 y 60 años ($M = 35.40$ años).

• PROCEDIMIENTO Y DISEÑO

La metodología de investigación fue de tipo experimental con un diseño intra-sujeto para contrastar la memoria de hechos vividos (autobiográfica) con la fabricada de hechos, basada en el visionado de un vídeo. Se manipularon dos condiciones experimentales. Primera, en la condición de memoria de hechos vividos (auto-experimentados, memoria autobiográfica) se pidió a los participantes que narraran la memoria de un hecho vivido (como ejemplos se les proporcionaron los referidos en el

DSM-5 como eventos traumáticos, tal como un la muerte o enfermedad de un familiar, operación quirúrgica o un accidente de tráfico) que cumpliera los criterios descritos por Steller (1989) para que fuera equiparable a la memoria obtenida en las evaluaciones forenses: implicación personal, tono emocional negativo del evento, y una gran pérdida de control. Segunda, en la condición de memoria fabricada de hechos se les presentó un material en vídeo de un robo solicitándoles que tomaran el testimonio como propio (experimentación directa) narrándolo como vivido (auto-experimentado). Además, los contenidos de lo visionado estaban fuera del control de los actores. De este modo se cumplían las condiciones (experimentación del evento como propio, carácter negativo del evento y situación fuera del control del observador) de una simulación de alta fidelidad (Steller, 1989).

Los participantes fueron recabados de la población general de modo aleatorio (disponibilidad y accesibilidad) preguntándoles si querían participar en un estudio de memoria. La participación fue voluntaria y sin ninguna contraprestación por ella. Los participantes pasaron por ambas condiciones. El orden de pase por las condiciones fue contrabalanceado (rotación estándar) a través de los participantes. La obtención de las memorias fue individual. Una vez recabadas las dos memorias (entrevistas) con cada sujeto se procedió a un debriefing en el que se les explicaba el objeto del estudio y se les pedía que informaran de la implicación y técnicas que habían usado para la creación del evento fabricado.

• MATERIAL E INSTRUMENTO DE MEDIDA

Como material del que desarrollar la memoria fabricada de hechos visionados se les proyectó a los participantes el vídeo de un robo en un domicilio (se verificó, mediante el consenso inter-investigadores que cumpliera las condiciones de una simulación de alta fidelidad) de una duración de 4 minutos y 30 segundos.

Las declaraciones de ambas condiciones, que fueron grabadas en vídeo, se transcribieron para el posterior análisis de contenido. Dos codificadores analizaron las declaraciones en los 19 criterios de realidad del CBCA (instrumento de medida): estructura lógica, elaboración inestructurada, cantidad de detalles, engranaje contextual, descripción de interacciones,

reproducción de conversaciones, complicaciones inesperadas durante el incidente, detalles inusuales, detalles superfluos, incomprensión de detalles relatados con precisión, asociaciones externas relacionadas, relatos del estado mental subjetivo, atribución al estado mental del autor del delito, correcciones espontáneas, admisión de falta de memoria, planteamiento de dudas sobre el propio testimonio auto-desaprobación, perdón al autor del delito, y detalles característicos del delito.

•FIABILIDAD DE LAS ENTREVISTAS Y CODIFICACIONES

Un entrevistador entrenado en la entrevista cognitiva, que favorece la manifestación de criterios de realidad en las declaraciones de hechos vividos (Vrij, 2008) y de indicios de engaño en las fabricadas (Vrij et al., 2008), obtuvo todas las declaraciones. Para el entrenamiento en esta modalidad de entrevista se siguió el programa descrito en Fishery Geiselman (1992). Esta entrevista se estructura en torno a 4 técnicas a emplear para la recuperación de la memoria (por este orden): reinstauración mental de contextos, narrar todo lo recordado, cambio de perspectiva y recuerdo en orden inverso. La fiabilidad de las entrevistas a los testigos es de tipo judicial, derivándose la falta de fiabilidad de la contaminación de la prueba dimanante de preguntas que puedan viciar la memoria del testigo (Pearse y Gudjonsson, 1996; Walsh y Milne, 2008) o de la aplicación de técnicas ilícitas de entrevista (Dando y Oxburg, 2016; May y Granhag, 2016), y científica, proveniente de la falta de producción de la misma. Ambos controles de fiabilidad recayeron en los investigadores que revisaron las entrevistas en busca de preguntas por parte del entrevistador u otras fuentes de contaminación (p.ej., interrupciones, toma de notas, no escucha activa) y el no seguimiento de las instrucciones de ejecución de la entrevista (los 4 pasos y por el orden prescrito). De observarse alguna fuente de contaminación, la entrevista (ampliada a las dos del participante por ser un diseño de medidas repetidas) sería eliminada del estudio (no se eliminó entrevista alguna). Asimismo, los investigadores revisaron todas las codificaciones de las entrevistas cotejando que todas tenían suficiente amplitud y para confirmar la cooperación del testigo en la recuperación de la memoria (productividad).

Se creó un manual para el entrenamiento en la codificación con las definiciones originales (Steller y Köhnken, 1989) de cada categoría y, en su caso, ejemplos de las mismas. Se diseñó un programa en línea con el descrito por Köhnken (2004) para el entrenamiento en el CBCA: Definición de los criterios y ejemplos —manual—; generación de nuevos ejemplos por los codificadores; transcripción de declaraciones; evaluación de material; y discusión de las evaluaciones discordantes y corrección de errores en la codificación. Se entrenaron a dos codificadores. Uno de ellos con experiencia previa y con práctica forense, y que ya se había mostrado consistente con otros codificadores en otros estudios (Arce et al., 2000). Como material de entrenamiento se les proporcionaron declaraciones de adultos, tanto de hechos vividos como fabricadas, que no formaron parte del presente estudio. Dichas declaraciones fueran previamente codificadas por investigadores y peritos. Cuando se observaba alguna discrepancia entre los codificadores de los evaluadores del estudio o con las codificaciones previas, se discutía con ellos sobre las mismas para corregir el error de codificación. Un codificador analizó todas las transcripciones de las declaraciones en los 19 criterios de realidad del CBCA. El segundo codificador con experiencia previa en la codificación con otros codificadores y también consistente consigo mismo en el tiempo (Fariña, Arce, Vilariño, y Novo, 2014), analizó el 20% de las transcripciones de las declaraciones, que fueron elegidas al azar del total, siendo la mitad de cada condición experimental. Como estadístico para la estimación de la concordancia inter-evaluadores (acuerdo) se tomó, al tratarse de una codificación en variables nominales con dos valores (presente vs. ausente, tal y como se codifican en la práctica forense), el índice kappa de Cohen (1960), que excluye de la concordancia la debida al azar. No obstante, este índice es incompleto si no se toman salvaguardas. Así y a modo de ejemplo que ambos codificadores registren en una declaración la categoría “complicaciones inesperadas durante el incidente” no implica directamente concordancia. De hecho, podrían estar refiriéndose a dos contingencias diferentes de modo que no se trataría de una concordancia (acierto) en la codificación, sino de dos errores. Para solventar este problema, Fariña et al. (2002) complementaron la codificación con la verificación de la exacta correspondencia

de las codificaciones. De este modo, se conoce la exactitud de la codificación y, por tanto, de la concordancia verdadera ($\bar{\kappa}$). Los resultados de la concordancia verdadera (ver Tabla 1) muestran una concordancia entre buena ($>.61$ — $<.81$) y muy buena ($\geq .81$).

En todo caso, esta medida de la concordancia no quiere decir que sea una medida realmente fiable (estimación de la variabilidad en la codificación introducida por los codificadores) ya que únicamente podemos concluir que esos dos codificadores son concordantes entre sí, en ese momento concreto y con ese material. Al respecto, Wicker (1975) estableció que para estimar la fiabilidad de las codificaciones no es suficiente con la concordancia inter-codificadores, sino también es preciso conocer la concordancia con

otros codificadores (terceros), intra-codificador (en el tiempo) e inter-contextos (i.e., codificación de material diferente al del estudio afectado por variables con efectos en la productividad, pertinencia y manifestaciones). Dado que uno de los codificadores fue consistente en otros estudios consigo mismo, con otros codificadores y en otros contextos, podemos sostener que la codificación por otros evaluadores con conocimientos y entrenamiento en la técnica resultaría en unos resultados concordantes. En suma, los datos son fiables (la variabilidad de la codificación estimada inter- e intra-evaluadores e inter-contextos entra en los límites de la aceptabilidad, κ), tal que otros codificadores entrenados en la técnica de análisis de contenido, aportarían resultados similares (fidelidad en la codificación).

Tabla 1
Concordancia verdadera ($\bar{\kappa}$) inter-codificador

CRITERIO	$\bar{\kappa}$
1. Estructura lógica	1.00
2. Elaboración inestructurada	.64
3. Cantidad de detalles	.85
4. Engranaje contextual	.87
5. Descripción de interacciones	.75
6. Reproducción conversaciones	1.00
7. Complicaciones inesperadas	1.00
8. Detalles inusuales	.87
9. Detalles superfluos	.86
10. Detalles incomprendidos relatados con precisión	1.00
11. Asociaciones externas relacionadas	.64
12. Estado mental subjetivo	1.00
13. Estado mental autor delito	1.00
14. Correcciones espontáneas	.87
15 Admisión falta de memoria	.87
16. Dudas propio testimonio	1.00
17. Auto-desaprobación	1.00
18. Perdón al autor del delito	1.00
19. Detalles característicos del delito	1.00

•ANÁLISIS DE DATOS

El análisis de los datos en términos de comparación de medias y de diferencias de proporciones es de gran utilidad científica y forense, pero de dichas comparaciones no se pueden extraer conclusiones robustas para la práctica profesional; en este caso, la forense. De ahí que sea necesario complementar los análisis de comparaciones con otros de estudio de casos que faciliten el desarrollo de técnicas de evaluación forense en diseños (casos) de $N = 1$ (American Psychiatric Association, 2014).

El estudio de la asociación entre los criterios de realidad (presente vs. ausente) y la memoria recabada (hechos vividos —autobiográfica— vs. hechos visionados —visionados en vídeo—) se abordó con la prueba de McNemar (muestras relacionadas) con la corrección de continuidad y con significación unilateral (la hipótesis Undeutsch sostiene que los criterios de realidad se encontrarán, en mayor medida —significativamente más— en declaraciones de memorias de hechos vividos). Este estadístico nos informa de la (in)dependencia de las variables que resulta necesario completar con la magnitud de tal relación. Para ello, se calculó la diferencia entre las proporciones observadas en ambas contingencias, $p_1 - p_2$, así como el intervalo de confianza para esa diferencia al 95% (si no incluye el cero es que es significativa, siendo el intervalo inferior la diferencia mínima a obtener) y el intervalo para la población al 80% (este intervalo incluye al 80% de los individuos de la población; si no incluye el cero se generaliza el efecto a cualquier contingencia, en tanto que si los incluye indica la existencia de moderadores en la dependencia de las variables). Para la estimación del tamaño del efecto en diferencia de proporciones se computó la h de Cohen (1988), equivalente a la d . Por su parte, el total de criterios de realidad para el factor tipo de memoria (auto-experimentada vs. no experimentada) fue analizado con la prueba t de student para muestras relacionadas y el tamaño del efecto con la d de Cohen (1988) para la que se obtuvo el intervalo de confianza al 95% (oscilación de la media en la población) y el intervalo para la población al 80% (en este intervalo comprende al 80% de los individuos de la población). Finalmente, y por su utilidad para la derivación de implicaciones para la práctica forense (Fariña, Redondo, Seijo, Novo,

y Arce, 2017), se han calculado los estadísticos probabilidad de superioridad (PS) y U de Cohen, en los análisis de comparación de medias. En las medidas en diferencias de proporciones, estos estadísticos no pueden obtenerse (la variable no cumple el supuesto de distribución normal) por lo que creamos un estimador ad hoc para el incremento de la probabilidad de correcta clasificación (ΔPCC) de casos. Este estadístico es la adaptación del BESD. Para ello transformamos h en r , ésta se interpreta en términos de incremento (si es positiva, y decremento si es negativa) en la probabilidad de correcta clasificación de casos en relación a la línea base/grupo control.

La magnitud del tamaño del efecto se interpretó en términos de las categorías tradicionales de Cohen (1988), esto es, pequeña ($d/h = 0.20$), moderado ($d/h = 0.50$), y grande ($d/h = 0.80$). A estas categorías se añadió la “más que grande” ($d/h = 1.20$), que crearon Arce, Fariña, Seijo y Novo (2015) para el contexto de la evaluación forense por las implicaciones en la interpretación de la clasificación de casos que trae aparejada. Además, esta categoría se corresponde con 1.20 DE dejando bajo de sí el 88.49% ($P_{88.49}$), frente al 57.92% de una magnitud pequeña, el 69.14% de la moderada y 78.81% de la grande. En la conversión de los tamaños del efecto se tomaron las tablas de Salgado (2018).

RESULTADOS

Los resultados de la comparación de los efectos en la manifestación de los criterios de realidad del factor tipo de declaración (hechos auto-experimentados vs. visionados en vídeo y no auto-experimentados) mostraron efectos significativos para el total de criterios de realidad del CBCA, $t(29) = 3.57$, $p < .001$, $d = 0.65$ con un IC al 95% [0.30, 1.0], es decir, el tamaño del efecto mínimo a obtener en la población (otros estudios en las mismas condiciones) es de 0.30. Ahora bien, este resultado no es generalizable a todas las memorias de los individuos de la población, IC para los individuos de la población al 80% [-3.44, 4.74]. En consecuencia, el total de criterios de realidad no discriminaría en todos los individuos entre memorias de hechos auto-experimentados y memorias fabricadas de hechos (visionados en vídeo). Al respecto, la probabilidad de superioridad (PS), esto es, de que una memoria de hechos auto-experimentados

Tabla 2
Capacidad discriminativa de los criterios realidad

Variable	χ^2	ρ	$\rho_1-\rho_2$	IC95%	$\rho_1-\rho_2$	IC80%	$\rho_1-\rho_2$	h	ΔPCC
CARACTERÍSTICAS GENERALES									
Estructura lógica	0.67	.125	.100	.024,	.176	-.284	.484	0.28	.139
Elaboración inestructurada	2.50	.055	.200	.099,	.301	-.312	.712	0.56	.270
Cantidad de detalles ¹	9.09	.000	.366	.244,	.488	-.251,	.983	1.31	.548
CONTENIDOS ESPECÍFICOS									
Engranaje contextual	4.00	.033	.233	.126,	.340	-.308	.774	0.62	.296
Descripción de interacciones	0.56	.227	-.135	-.221,	-.049	-.572	.302	-0.30	-.148
Reproducción de conversaciones	3.50	.029	.267	.155,	.379	-.299	.833	0.66	.313
Complicaciones inesperadas	0.67	.125	-.100	-.176,	-.024	-.484,	.284	-0.28	-.139
PECULIARIDADES DEL CONTENIDO									
Detalles inusuales	0.07	.396	.067	.004,	.130	-.253,	.387	0.17	.085
Detalles superfluos	2.08	.073	.200	.099,	.301	-.312,	.712	0.52	.252
Incomprensión detalles relacionados	----	----	----	----	----	----	----	----	----
Asociaciones externas relacionadas	0.50	.500	.000	.000,	.000	.000,	000	0.00	.000
Relatos estado mental subjetivo	0.08	.500	.000	.000,	.000	.000,	000	0.00	.000
Atribución estado mental agresor	4.00	.020	-.234	-.341,	-.127	-.311,	.779	-0.75	.351
CONTENIDOS REFERENTES A LA MOTIVACIÓN									
Correcciones espontáneas	2.08	.073	.200	.099,	.301	-.312,	.712	0.52	.252
Admisión de falta de memoria	1.07	.151	.167	.073,	.261	-.310,	.644	0.37	.182
Plantear dudas sobre testimonio	----	----	----	----	----	----	----	----	----
Auto-desaprobación ¹	4.16	.016	.200	.099,	.301	-.312,	.712	0.93	.422
Perdón al autor del delito	0.00	.500	-.033	-.078,	.012	-.212,	.146	-0.35	-.172
ELEMENTOS ESPECÍFICOS DE LA AGRESIÓN									
Detalles característicos del delito	----	----	----	----	----	----	----	----	----

Nota: $g(1)$; prueba de McNemar con corrección de continuidad para muestras pequeñas, significación unilateral; ¹No se registró esta contingencia en memorias fabricadas de hechos.

contenga más criterios de realidad que una de fabricados, es de .68; la distribución de las áreas de las declaraciones de hechos vividos y fabricada de hechos ($U_1 = .41$) no se superponen en el 41%, es decir, son totalmente independientes; el 63% de las memorias fabricadas de hechos ($U_2 = .63$) contienen más criterios de realidad que el 63% con menos criterios de realidad de las memorias de hechos vividos, es decir, la tasa de clasificación de las memorias fabricadas como de hechos auto-experimentados sería de .63 (un 13% sobre el azar, 50%) y que la media de criterios de realidad en las memorias auto-experimentadas se corresponde con el percentil 74 ($U_3 = .74$) en las memorias fabricadas de hechos, es decir, la tasa de correcta clasificación de las memorias de hechos vividos sería del 74% (24% sobre el azar).

En relación a los criterios de realidad (ver Tabla 2), los resultados informan que los criterios elaboración inestructurada, cantidad de detalles, engranaje contextual, reproducción de conversaciones, detalles superfluos, correcciones espontáneas y auto-desaprobación se han registrado significativamente más (un 20, 36.6, 23.3, 26.7, 20, 20 y 20%, respectivamente) en las memorias de declaraciones de hechos auto-experimentados. Estos resultados son generalizables a otros estudios con la misma población y condiciones (el intervalo al 95% para la diferencia de proporciones a obtener en la población no incluye el cero, es decir, la proporción de memorias de hechos vividos en las que se registre el criterio siempre será superior a la proporción a registrar en la memoria fabricada de hechos), oscilando entre el 9.9 y el 30.1% en el criterio elaboración inestructurada; entre el .244 y .488 en cantidad de detalles; entre .126 y .340 en engranaje contextual; entre .155 y .379 en reproducción de conversaciones; entre .099 y .301 en detalles superfluos; correcciones espontáneas y auto-desaprobación. La magnitud del efecto fluctúa entre moderada ($h = 0.50$) para elaboración inestructurada, engranaje contextual, reproducción de conversaciones, detalles superfluos y correcciones espontáneas; grande ($h = 0.80$) en auto-desaprobación; y más que grande ($h = 1.20 = 1.20 DT = P88.49$, es decir, el 88.49% de los tamaños del efecto son más pequeños) en cantidad de detalles. No obstante, estos resultados no son generalizables a diseños de $N = 1$ (el intervalo al 80% para la diferencia de proporciones en los

sujetos, incluye en cero). En términos de utilidad práctica, el incremento en la probabilidad de correcta clasificación (ΔPCC) de las memorias de hechos auto-experimentados asociados a los criterios elaboración inestructurada, cantidad de detalles, engranaje contextual, reproducción de conversaciones, detalles superfluos, correcciones espontáneas y auto-desaprobación es .270, .548, .296, .313, .252, .252 y .422, respectivamente.

Por su parte, el criterio de realidad atribución al estado mental del autor del delito se ha registrado significativamente menos, el 23.4%, en las declaraciones de hechos vividos que en las fabricadas de hechos (resultado contrario a la hipótesis de realidad de Undeutsch). Asimismo, este resultado es generalizable a la población de referencia y en las condiciones del diseño (invariable). No obstante, no se puede generalizar a todos los sujetos (puede registrarse en sujetos en declaraciones de hechos vividos).

A su vez, los criterios estructura lógica, descripción de interacciones, complicaciones inesperadas durante el incidente, detalles superfluos, asociaciones externas relacionadas, relatos del estado mental subjetivo, admisión de falta de memoria y perdón al autor del delito no discriminan entre memorias de hechos auto-experimentados y memorias no experimentadas (fabricadas basadas en hechos visionados); es decir, no confirman la hipótesis.

Finalmente, destaca que los criterios de realidad incomprensión de detalles relatados con precisión, planteamiento de dudas sobre el testimonio y detalles característicos del delito no se mostraron productivos en ninguna de las dos condiciones.

DISCUSIÓN

Los resultados previos están sujetos a limitaciones que se han de tener en cuenta a la hora de concretar su alcance. Primero, la situación no es real por lo que los participantes no actúan como en tal circunstancia (implicación), con efectos directos en los resultados (validez aparente) en la investigación en psicología forense (Fariña, Real, y Arce, 1994). Al respecto, la literatura ha constatado sistemáticamente (ver revisiones meta-analíticas de menores y adultos; Amado et al., 2015, 2016; Oberlader et al., 2016) que la capacidad discriminativa y de clasificación de memorias de hechos auto-experimentados de los

criterios de realidad del CBCA es significativamente mayor en estudios de campo (casos reales) que en estudios de laboratorio (casos simulados). Segundo, la longitud de la declaración tiene efectos en la productividad de las categorías de contenido (Arce, 2017; Köhnken, 2004), siendo significativamente menor en las memorias autobiográficas que en las reales de eventos de victimización de un delito, y las fabricadas en condiciones experimentales de las fabricadas con propósitos judiciales. Tercero, si bien para las memorias autobiográficas se concretó (y verificó) que cumplieran unas condiciones a fin de asemejarlas a memorias reales de victimización (experimentadas de firma directa, de carácter negativo y que la propia situación y resultado se encuentren fuera del control de la persona) no son intercambiables con memorias reales (Gödert, Gamer, Rill, y Vossel, 2005). A modo de ejemplo, la categoría perdón al autor del delito sólo puede resultar productiva en las memorias reales si se describe un delito, en tanto en memorias reales de victimización delictiva esta contingencia es sistemáticamente factible. Cuarto, asimismo las memorias fabricadas de hechos se ciñen al evento visionado, tal que las características específicas del evento y del sujeto pueden generar una producción dispar en las categorías de contenido (Steller, 1989; Undeutsch, 1989). Quinto, la potencia estadística de los diseños es adecuada para tamaños moderados, grandes y más que grandes del efecto, pero escasa para tamaños pequeños. Sobre la base de estas limitaciones de los resultados del estudio se derivan las siguientes conclusiones:

1) Los resultados avalan la viabilidad y robustez, en general, de la hipótesis Undeutsch para la discriminación entre memorias de hechos auto-experimentados y memorias fabricadas de hechos basadas en el visionado de un vídeo. Así, el total de criterios de realidad del CBCA discrimina significativamente entre ambas memorias en línea con dicha hipótesis (más criterios de realidad en las declaraciones de hechos vividos). Dado que en la práctica forense no es admisible la clasificación de memorias fabricadas como de hechos vividos (vulnera el principio de presunción de inocencia al incriminar a un inocente), se ha de formular un criterio de decisión estricto que reduzca a cero tal probabilidad (Arce, 2017), no gozando de validez forense ni las reglas de decisión (p.ej.,

las memorias de hechos auto-experimentados contienen más criterios de realidad o se manifiestan más intensamente; Undeutsch, 1989), ni la clasificación de la memoria en una escala Likert de decisión sobre la credibilidad del testimonio (creíble, probablemente creíble, indeterminado, probablemente increíble, no creíble; Steller, 1989), las clasificaciones intuitivas (Steller y Köhnken, 1989) o basadas en un juicio clínico (Köhnken, 2004).

2) Los criterios de realidad elaboración inestructurada, cantidad de detalles, engranaje contextual, reproducción de conversaciones, detalles superfluos, correcciones espontáneas y auto-desaprobación se han registrado significativamente más en las declaraciones de hechos vividos. Especialmente reseñable es que el criterio auto-desaprobación que había resultado no significativo (esto es, si bien el tamaño del efecto promedio era positivo, el intervalo de confianza para dicho efecto incluía el 0 de modo que en estudios se pueden obtener resultados contrarios a la hipótesis) en la discriminación entre memorias de eventos auto-experimentados y fabricadas en población adulta en el meta-análisis de Amado et al. (2016), este resultado avala que este criterio no sólo es efectivo en dicha discriminación ($h = 0.93$), sino altamente robusto en la correcta clasificación de las memorias auto-experimentadas en relación a las fabricadas visionadas (línea base/grupo control) en vídeo ($\Delta PCC = .422$).

3) El criterio atribución al estado mental del autor de delito sea registrado significativamente más en declaraciones de hechos fabricados visionados en vídeo. Del mismo modo, clasifica, contrariamente a la hipótesis Undeutsch, las declaraciones fabricadas como vividas. Este resultado, además, sería generalizable a la población, aunque no a todas memorias fabricadas de hechos visionados. Ahora bien, este resultado es muy probable que esté mediado por una de las limitaciones del diseño del estudio (Herbón, Novo, y Amado, 2016): las memorias autobiográficas no tuvieron por objeto la descripción de victimización delictiva con un enfrentamiento directo con un agresor (sólo se esta contingencia observó en dos testimonios). Por tanto, este resultado no ha de tomarse como válido.

4) Las categorías estructura lógica, descripción de interacciones, complicaciones inesperadas, detalles inusuales, asociaciones externas relacionadas, relatos del estado mental subjetivo, admisión de falta de memoria y perdón al autor del delito, no discernen entre memorias de hechos auto-experimentados y memorias fabricadas de hechos visionados (no experimentados). Resultado generalizable a la población de declaraciones, pero no a todos los individuos. En todo caso, estos resultados no implican que, en otros contextos, esto es, eventos con otras características (limitación del diseño), la capacidad discriminatoria y de clasificación entre memorias fabricadas de hechos (p.ej., agresiones sexuales) y memorias de hechos vividos (p.ej., que supongan la victimización de un delito) sea otra.

5) Las categorías incompreensión de detalles relatados con precisión, planteamiento de dudas sobre el testimonio y detalles característicos de la agresión no resultaron productivas. No obstante, de esto no se puede concluir que sea generalizable a memorias fabricadas de eventos de otra naturaleza (características del evento) y memorias de victimización delictiva y con diferentes características (limitaciones del diseño). De hecho, una revisión meta-analítica encontró que estas categorías discriminaron significativamente entre memorias de hechos auto-experimentados y fabricadas (no visionadas) en adultos (Amado et al., 2016).

6) En términos de clasificación de casos, hallamos (ténganse en cuenta las limitaciones del diseño en el alcance de este hallazgo) que los criterios de realidad (algunos) discriminan entre memorias de hechos vividos y fabricadas de hechos visionados en vídeo (el criterio atribución al estado mental del autor del delito encontramos el efecto contrario, pero ya advertimos de su falta de validez), pero ninguno de ellos es totalmente sensible (no clasifica todas las memorias de hechos vividos como reales) y específico (no clasifica todas memorias fabricadas como tales) en la clasificación de la realidad de la memoria. En consecuencia, si bien ciertos criterios de realidad clasifican significativa y correctamente las memorias de hechos auto-experimentados, no son prueba suficiente (no clasifica correctamente todas las memorias reales) y válida (clasifica memorias fabricadas como reales)

por sí mismos para la tarea forense. Asimismo, el total de criterios clasifica correctamente las memorias de hechos vividos (verdaderos positivos) en el 74% de los casos (el 26% restante no podría ser clasificado como memorias reales, tampoco como memorias fabricadas pues la técnica sólo es aplicable para la clasificación de la realidad de la memoria). Ahora bien, del total de criterios de realidad del CBCA también se desprende que la probabilidad de clasificación de las memorias fabricadas como reales (falsos positivos) es de .63. En consecuencia, el total de criterios de realidad no es un indicador válido (en la práctica forense no es admisible el falso positivo porque supone la incriminación/condena de un inocente) en términos forenses para la clasificación de las memorias reales. Esta conclusión también se ha encontrado en todo tipo de memorias de eventos (características del evento), de poblaciones (menores, adultos, testigos, encausados) y de diseños (experimentales, de campo) (Amado et al., 2015, 2016; Oberlader et al., 2016).

En conclusión, si bien la hipótesis Undeutsch es válida para discernir (y clasificar) entre memorias de hechos auto-experimentados y fabricadas de hechos (Amado et al., 2015, 2016; Oberlader et al., 2016), incluidas aquellas basadas en hechos visionados en vídeo (este estudio), la técnica derivada de la misma de referencia, el SVA, que incluye el análisis de contenido basado en los criterios de realidad del CBCA, no cumple con el criterio (Arce, 2017)

Futura investigación debería profundizar más en la capacidad discriminativa y de clasificación de los criterios de realidad entre memorias de hechos auto-experimentados y fabricadas de hechos basadas en el visionado de los mismos (no experimentados), especialmente abordándola con diseños de investigación que confronten las limitaciones apuntadas al presente estudio.

• **Conflicto de intereses**

Los autores declaran no tener ningún conflicto de intereses.

REFERENCES

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.

- <https://doi.org/10.1016/j.ejpal.2014.11.002>
- Amado, B. G., Arce, R., Fariña, F., y Vilariño, M. (2016). Criteria-Based Content Analysis (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>
- American Psychiatric Association. (2014). *Manual diagnóstico y estadístico de los trastornos mentales (DSM-V)*. Madrid, España: Panamericana.
- 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, 171-190. <https://doi.org/10.5944/ap.14.2.21347>
- Arce, R., Fariña, F., y Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema*, 12, 623-628. Recuperado de <http://www.psicothema.com/pdf/381.pdf>
- Arce, R., Fariña, F., Seijo, D., y Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment*, 22, 769-777. <https://dx.doi.org/10.1177/1073191114558111>
- Boccaccini, M. T. (2002). What do we really know about witness preparation? *Behavioral Sciences & the Law*, 20(12), 161-189. <https://doi.org/10.1002/bsl.472>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2a. ed.). Hillsdale, NJ: LEA.
- Dando, C. J., y Oxburgh, G. E. (2016). Empathy in the field: Towards a taxonomy of empathic communication in information gathering interviews with suspected sex offenders. *European Journal of Psychology Applied to Legal Context*, 8, 27-33. <https://doi.org/10.1016/j.ejpal.2015.10.001>
- Fariña, F., Arce, R., y Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema*, 14, 39-46. Recuperado de <http://www.psicothema.com/pdf/684.pdf>
- Fariña, F., Arce, R., Vilariño, M., y Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *The Spanish Journal of Psychology*, 17, e32, 1-10. <https://dx.doi.org/10.1017/sjp.2014.30>
- Fariña, F., Arce, R., y Real, S. (1994). Ruedas de identificación: De la simulación y la realidad. *Psicothema*, 6, 395-402. Recuperado de <http://www.psicothema.com/psicothema.asp?id=935>
- Fariña, F., Redondo, L., Seijo, D., Novo, M., y Arce, R. (2017). A meta-analytic review of the MMPI validity scales and indexes to detect defensiveness in custody evaluations. *International Journal of Clinical and Health Psychology*, 17, 128-138. <https://doi.org/10.1016/j.ijchp.2017.02.002>
- Fisher, R. P., y Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.
- Gödert, H. W., Gamer, M., Rill, H. G., y Vossel, G. (2005). Statement Validity Assessment: Inter rater reliability of Criteria Based Content Analysis in the mock crime paradigm. *Legal and Criminological Psychology*, 10, 225-245. <https://doi.org/10.1348/135532505X52680>
- Kaplan, M. F., y Anderson, N. H. (1973). Information integration theory and reinforcement theory as approaches to interpersonal attraction. *Journal of Personality and Social Psychology*, 28, 301-312. <https://dx.doi.org/10.1037/h0035112>
- 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>
- Lee, Z., Klaver, J. R., y Hart, S. D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime & Law*, 14, 73-84. <https://dx.doi.org/10.1080/10683160701423738>
- May, L., y Granhag, P. A. (2016). Using the Scharff-

- technique to elicit information: How to effectively establish the "illusion of knowing it all"? *European Journal of Psychology Applied to Legal Context*, 8, 79-85. <https://doi.org/10.1016/j.ejpal.2016.02.001>
- Novo, M., Herbón, J., y Amado, B. G. (2016). Género y victimización: Efectos en la evaluación de la violencia psicológica sutil y manifiesta, apego adulto y tácticas de resolución de conflictos [Victimization and gender: Effects in the evaluation of subtle and overt violence, adult attachment and conflict resolution tactics]. *Revista Iberoamericana de Psicología y Salud*, 7, 89-97. <https://doi.org/10.1016/j.rips.2016.05.002>
- 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://sepjif.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=26&Itemid=110&lang=en
- Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., y Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior*, 40, 440. <https://dx.doi.org/10.1037/lhb0000193>
- Ostrom, T. M., Werner, C., y Saks, M. J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 36, 436-450. <http://dx.doi.org/10.1037/0022-3514.36.4.436>
- 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>
- Salgado, J. F. (2018). Transforming the Area under the Normal Curve (AUC) into Cohen's d, Pearson's r pb, Odds-Ratio, and Natural Log Odds-Ratio: Two Conversion Tables. *European Journal of Psychology Applied to Legal Context*, 10, 35-47. <https://doi.org/10.5093/ejpalc2018a5>
- 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. 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.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen [La evaluación de la credibilidad de los testigos]. 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.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2a. ed.). Chichester, UK: John Wiley and Sons.
- Vrij, A., Mann, S. A., Fisher, R. P., 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. <https://doi.org/10.1007/s10979-007-9103-y>
- 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>
- Wicker, A. W. (1975). An application of the multitrait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 1, 575-579. <https://dx.doi.org/10.1177/014616727500100405>