



Implanting rich autobiographical false memories: Meta-analysis for forensic practice and judicial judgment making

Ramón Arce^{a,*}, Adriana Selaya^a, Jéssica Sanmarco^a, Francisca Fariña^b

^a Unidad de Psicología Forense, Facultad de Psicología, Universidade de Santiago de Compostela, Spain

^b Departamento AIPSE, Universidade de Vigo, Spain



ARTICLE INFO

Keywords:

Suggested memories
Fabricated memories
Expert testimony
Memory wars
Recovered memory

ABSTRACT

Objective: The implanting of rich autobiographical false memories is crucial for judicial decision-making, and the forensic evaluation of a testimony. In order to assess this issue, a meta-analytical review of the probability of implanting rich autobiographical false memories was performed.

Method: A total of 30 primary studies analysing the probability of implanting rich autobiographical false memories were retrieved. Random-effects meta-analyses correcting the effect size for sampling error were performed.

Results: The results revealed a significant, positive, generalizable (the lower limit for the 80% credibility value was $d = 1.13$), and more than large mean effect size ($d = 1.43[1.33, 1.53]$) for the implanting of false memory. The moderating effects of stimulus type showed that the effect of the probability of implanted false memory was significantly higher in experienced events ($d = 2.03[1.63, 2.43]$) than in false narratives ($d = 1.35[1.23, 1.47]$), and in doctored photographs ($d = 1.29[1.06, 1.52]$). A similar effect for memory implantation was observed in both the underage ($d = 1.44[1.29, 1.59]$), and in adults ($d = 1.36[1.22, 1.50]$). The moderator techniques for implanting false memories revealed a significantly lower probability of implanting false rich memory with non-directive instructions ($d = 0.90[0.53, 1.27]$) than with guided imagery ($d = 1.45[1.32, 1.58]$), or with pressure to answer ($d = 1.56[1.17, 1.95]$) instructions. The event emotional valence moderator exhibited the same effect for positive ($d = 1.27[1.09, 1.45]$) and negative valence events ($d = 1.30[1.17, 1.43]$).

Conclusions: The implications of the results for forensic testimony evaluation, police interrogations, and judicial cross-examination are discussed.

Introduction

The system of justice is the public means of conflict resolution, and of protecting legal assets, both of which are necessary and useful conditions for the development of individual and social life. The procedural activity underpinning the justice system is the evidence, and the burden of proof lies with the prosecution (criminal justice). The accuser's testimony constitutes the main evidence of a crime committed in the private sphere such as sexual abuse, sexual assault, or family violence (Novo & Seijo, 2010). Memory contains a register of individual experiences that is the corner stone of a testimony. However, a witness' memory in the courtroom is not sufficient evidence in itself, and must be submitted to cross-examination and/or forensic evaluation to estimate both its reliability (credibility), and validity (the value of the testimony as evidence for judgment making). In judicial settings, witnesses are classified as either credible (with value as evidence), or not credible (without value as evidence), which correspond to the honest or dishonest witness in the psychology of testimonies. Nevertheless, research on the memory of witnesses has found the accounts of honest witnesses contained inaccurate

information (Köhnken et al., 1999; Memon et al., 1997, 2010). Hence, and in line with the meta-analytical reviews (Memon et al., 2010), it has been estimated that 10 to 20% of the information of an honest witness is inaccurate (Arce, 2017). The lack of accurate information reported by witnesses is the object of judicial evaluation through cross-examination. As a result of repeated cross examination, inaccuracies may arise in honest testimonies owing to inexact memory, and the interrogation techniques. Consequently, the lack of internal consistency of the testimony through time, and the addition of other evidence may undermine or annul the reliability (credibility), and in turn, the validity of the testimony (without value as evidence) in judicial judgment making (Arce, 2017). Thus, in order to bestow credibility to a complainant witness' testimony, forensic evaluation techniques involving the content analysis of statements have been designed. These techniques do not validate exactitude, but rather the external origin of memory (Gancedo et al., 2021; Sporer & Küpper, 2004), reality memory (Undeutsch, 1967; Steller & Köhnken, 1989; Vrij et al., 2021), or memory of self-experienced events (Arce & Fariña, 2014). Nonetheless, cognitive researchers have pointed out that an honest testimony (statement, account) can be

* Corresponding author at: Facultad de Psicología, Campus Vida, s/n, 15782 Santiago de Compostela, Spain.

E-mail address: ramon.arce@usc.es (R. Arce).

the product of false memories implanted by using suggestive techniques employed in several psychotherapeutic practices (Lindsay & Read, 1994). The identification of patient memory reports of abusive experiences in psychotherapy as false memories by cognitive researchers has led to the 'memory wars' in science (Loftus & Ketcham, 1994), which have become among the most controversial debates in the history of psychology (Davis & Loftus, 2015), and continue to be contentious (Pathiis et al., 2014). However, false memories in the forensic field are not only associated to suggestive techniques in psychotherapy, but also fundamentally to both police and judicial interviews (Gudjonsson, 2003; Memon et al., 2010; Deeb et al., 2022). That is, how the initial statement was obtained, the subsequent interviews carried out for judicial proceedings, extrajudicial interviews, and forensic techniques employing suggestive techniques to obtain a testimony (e.g., anatomical dolls, interrogations) (Arce, 2017; Köhnken, 2004).

So, what are false memories? The best definition is the operative one i.e., the one used by memory studies measuring false memories. There are three measurement paradigms: DRM lists, misinformation effect, and memory implantation. The DRM lists (Deese-Roediger-Mcdermott word lists; Deese, 1959; Roediger & McDermott, 1995) involve participants reading a list of semantically related words which evoke a 'critical lure'. After a delay, participants are encouraged to recognize the list of words introducing a non-presented semantically related word. The recognition or recall of the critical word is the measure of the false memories. Indeed, this is not a false memory in terms of the statement of a complainant (an event, account), in judicial judgment making this effect is a suggested error. The misinformation effect paradigm consists of presenting an event to participants (usually using a video, slide or photo) with the later presentation of narratives or questions that include misleading information about the presented material (Loftus et al., 1978; Zaragoza & Mitchell, 1996). False memories are created via the introduction of the misleading information (ordinarily through questions or a narrative descriptions) as part of a 'complex' event (for example, a slide of a car stopped at a stop sign; nevertheless, these are not complex in terms of the judicial memory assessed in forensic reports, which exceeds the witness' learning ability; Arce, 2017; Köhnken, 2004; Undeutsch, 1968). Finally, the memory of participants is measured with false memories being the misleading information provided (for example, a yield sign instead of a stop sign). Once again, this is not a false memory in terms of a statement of a complainant. In fact, the material is from non-self-experienced memories, not self-experienced memories (autobiographical), and the implanted memory consists of a distortion of memory measuring if the implanted inaccurate information becomes part of the original memory, not in the implanting of a non-experienced event (or a partial, complete in itself and complex event). A third research paradigm referred to as 'memory implantation' (Loftus & Prickell, 1995; Lindsay et al., 2004; Wade et al., 2002) is based on suggestive techniques to implant the memory of a pseudoevent or a real event not remembered by the participant. The suggestive techniques employed encompass guided imagery (interviewers guide participants to imagine events as if they really had happened; Lindsay et al., 2004; Wade et al., 2002); prompts/cues (interviewers provide prompts/cues to participants to remember a non-recalled event; Ost et al., 2005); pressure to answer (interviewers urge participants to recall because their friends, parents or others remembered or already told the event; Bruck et al., 2002; Principe et al., 2013); and a non-directive technique (the interviewer asks the participant to think about the event; Loftus et al., 1995).

The measure of the effect (false memory) is through the recall of the implanted event (rich memory). This task (implanting a memory) and the measure of the effects (entire events) are close to real life situations of the memory reported by claimants in court, and in forensic assessment (Otgaar & Candel, 2011). Nonetheless, the implantation of false memories has not been the main focus of the laboratory research. Thus, a bibliometric study found that the interest in the study of implanting rich false memories (implantation paradigm: memories of entire events) was less, 13.1%, than in the implanting or recognition of details

(misinformation paradigm: memory of details –non-rich false memories), 31.9% (Pezdek & Lam, 2007).

The present study

These findings, and the requirements for the forensic evaluation of testimonies underscore the need to conclusively determine whether it is possible to implant a complex event in a witness' memory. This implies forensic testimony evaluation techniques must be able to discriminate not only between the memory of honest versus dishonest testimonies (intentionally fabricated memories of an account), but also between honest testimonies of self-experienced memories of events from rich implanted memories of events. Thus, a meta-analytical review of the literature was performed to examine the claim of cognitive researchers that rich false memories of an event can be implanted, to quantify the prevalence, and to evaluate the effects of moderators relevant to forensic evaluation.

Method

Database search of studies

The initial search for studies was a sensitive one performed with 'false memory' as descriptor in the scientific reference databases Web of Science (all fields), Scopus (article title, abstract, keywords) and PsycInfo (any field). The search returned 4545 primary studies in the Web of Science, 7358 in Scopus, and 3112 in PsycInfo, which were reduced to 4255 studies after eliminating duplicates and studies that did not implant false memories. In order to avoid a gray literature bias or leaving significant studies out, a 'snowball method' was applied to the located studies that consisted in reviewing the references of the selected studies. This procedure identified 202 additional papers. All studies were screened if the formation/induction of false memories (focus of the review) had been evaluated, with a total of 1381 studies being selected (see in Table 1 the study characteristics).

No language, population, methodology, or publication time restriction was applied. The last search date was December 2022.

Quantitative journal article reporting standards (JARS), and PRISMA guidelines were followed, if possible, in the preparation of the manuscript.

Inclusion and exclusion criteria

The following inclusion criteria (requirements for the validity of the results for forensic practice, judicial judgment making or the performance of a meta-analysis) were applied to the results of the database searches: a) studies implanting rich autobiographical false memories (implantation paradigm: implantation of a memory of an entire event); b) studies assessing false memories using free recall; and c) studies reporting the effect size for memory implantation or data to calculate it. The exclusion criteria, based on the lack of validity for forensic practice and/or judicial judgment making, were: a) studies using an alternative method to rich false memories to implant false memories (implantation of details, not of an entire event; misinformation paradigm, DRM lists); b) research studies of false memories from a physiological approach (these did not measure memory but physiological registers); c) studies which implanted false memories from films or videos (not autobiographical memory); d) studies assessing source monitoring errors or using misleading questions (memory of details, not entire memories of an event); and e) crashing memories or social contagion studies in which the false memory was not autobiographical.

After the application of the inclusion and exclusion criteria, a total of 30 primary studies were included in the meta-analysis (see flow diagram in Fig. 1), and a total of 38 effect sizes were obtained.

Table 1
Primary studies data.

Study Cite	Total N	Moderator	<i>h</i> [95% CI]
Bruck et al. (2002) (Negative event)	16	False narrative Underage (< 13 y/o) Pressure to answer Negative event	2.25[2.06, 2.44]
Bruck et al. (Positive event)	16	False narrative Underage (< 13 y/o) Pressure to answer Positive event	2.09[1.88, 2.30]
Calado et al. (2021) (Standard implantation)	39	False narrative Adults Guided imagery Negative event	0.94[0.81, 1.07]
Calado et al. (2021) (Repeated implantation)	41	False narrative Adults Guided imagery Negative event	0.97[0.84, 1.10]
Calado et al. (2021) (Gradual implantation)	38	False narrative Adults Guided imagery Negative event	1.25[1.10, 1.40]
Desjardins and Scoboria (2007)	44	False narrative Adults Guided imagery Negative event	1.62[1.47, 1.77]
Garry and Wade (2005)	44	Adults Guided imagery Positive event	1.89[1.75, 2.04]
Garry and Wade (2005)	22	Doctored photograph	1.57[1.36, 1.78]
Garry and Wade (2005)	22	False narrative	2.26[2.10, 2.42]
Heaps and Nash (2001)	63	False narrative Adults Guided imagery Doctored photograph Adults Guided imagery Positive event	1.30[1.18, 1.42]
Hessen-Kayfitz and Scoboria (2011)	82	Doctored photograph Adults Guided imagery Positive event	1.25[1.15, 1.35]
Huffman et al. (1997)	22	False narrative Underage (< 13 y/o) Pressure to answer	0.76[0.62, 0.90]
Hyman and Billings (1998)	66	False narrative Negative event	1.10[0.99, 1.21]
Hyman et al. (1995) (I)	20	False narrative Adults Prompts/Cues	0.93[0.75, 1.11]
Hyman et al. (1995) (Experiment I)	20	Positive event	0.64[0.51, 0.77]
Hyman et al. (1995) (Experiment I)	20	Negative event	0.64[0.51, 0.77]
Hyman et al. (1995) (Experiment II)	51	False narrative Adults Prompts/Cues Negative event	1.06[0.94, 1.18]
Hyman and Pentland (1996)	65	False narrative Adults Negative event	1.04[0.94, 1.14]
Hyman and Pentland (1996)	32	Guided imagery	1.32[1.15, 1.49]
Hyman and Pentland (1996)	33	Non-directive technique	0.71[0.60, 0.82]
Lindsay et al. (2004)	45	False narrative Adults Guided imagery Negative event	1.82[1.68, 1.96]
Loftus and Pickrell (1995)	24	False narrative Adults Non-directive technique Negative event	1.05[0.88, 1.22]
Mazzoni and Memon (2003)	72	False narrative Adults Guided imagery Negative event	1.85[1.74, 1.96]
Ost et al. (2005)	31	False narrative Adults Prompts/Cues Negative event	0.99[0.84, 1.14]
Ost et al. (2005)	31	Positive event	0.73[0.61, 0.85]

(continued)

Table 1 (Continued)

Study Cite	Total N	Moderator	h[95% CI]
Otgaar et al. (2008)	76	False narrative Underage (< 13 y/o) Guided imagery	2.06[1.96, 2.16]
Otgaar et al. (2009)	91	Negative event False narrative Underage (< 13 y/o) Guided imagery	1.43[1.33, 1.53]
Otgaar et al. (2010)	120	Negative event False narrative Underage (< 13 y/o) Guided imagery	1.44[1.35, 1.53]
Otgaar et al. (2012)	45	Negative event False narrative Underage (< 13 y/o) Guided imagery	1.59[1.44, 1.74]
Pezdek et al. (1997) (Experiment I)	51	Positive event False narrative Underage (13–18 y/o) Non-directive technique	1.06[0.94, 1.18]
Pezdek et al. (1997) (Experiment II)	20	Negative event False narrative Non-directive technique	0.80[0.64, 0.96]
Pezdek and Hodge (1999)	39	Negative event False narrative Underage (< 13 y/o)	1.49[1.33, 1.65]
Porter et al. (1999)	77	Pressure to answer Negative event False narrative Adults	1.69[1.58, 1.80]
Principe et al. (2013)	40	Guided imagery Negative event Experienced event Underage (< 13 y/o)	1.77[1.62, 1.92]
Principe et al. (2017)	108	Pressure to answer Experienced event Underage (< 13 y/o) Prompts/Cues	2.16[2.08, 2.24]
Quas et al. (1999)	40	Prompts/Cues False narrative Underage (< 13 y/o) Prompts/Cues	1.42[1.27, 1.57]
Shaw and Porter (2018)	60	Negative event False narrative Adults Guided imagery	2.06[1.95, 2.17]
Strange et al. (2006)	115	Negative event Doctored photograph Underage (< 13 y/o) Prompts/Cues	1.13[1.05, 1.21]
Strange, Wade and Hayne (2008)	105	Positive event False narrative Adults Prompts/Cues	1.11[1.02, 1.20]
Strange, Hayne and Garry (2008)	62	Positive event Doctored photograph Underage (< 13 y/o) Prompts/Cues	1.24[1.12, 1.36]
Wade et al. (2002)	20	Positive event Doctored photograph Adults Guided imagery	1.57[1.35, 1.79]
Wade et al. (2010)	53	Positive event Doctored photograph Adults Guided imagery Positive event	1.70[1.57, 1.83]

Note. h: effect size h[95% CI].

Coding of primary studies

The following data from the studies were coded for the meta-analysis: a) primary study reference; b) document type (article, doctoral thesis, proceeding paper, chapter); c) sample characteristics (i.e., age, gender, sample size); d) estimation of the reliability/concordance of codings; and e) reported effect size for implanting memory, or the data

necessary to compute it. Additionally, a successive approach procedure was employed for the definition of the encoding categories of moderators relevant for forensic evaluation (Vilariño et al., 2013). This consisted of two researchers with knowledge in scanning all the selected papers to search for moderators. Identified variables were discussed and a consensus on moderators was reached by the researchers. Thus, four moderators of relevance for a testimony were coded: a) the type of

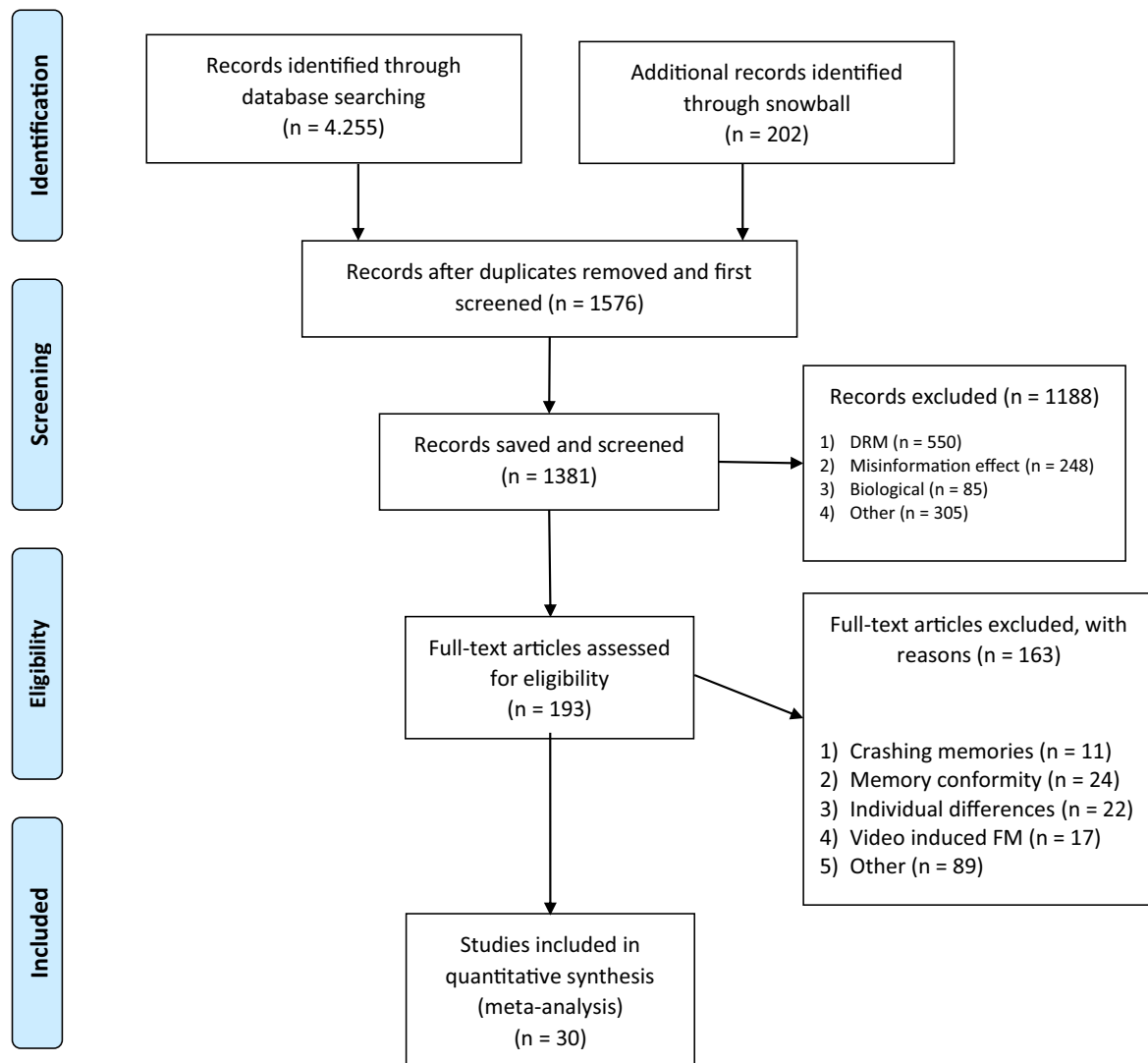


Fig. 1. Flow diagram of the meta-analysis.

stimuli (materials employed for memory implantation); b) the age of the participants (underage and adults); c) the suggestive technique employed for implanting false memories; and d) the emotional valence of the event (positive and negative events).

Two experienced and trained raters analysed independently the studies in these categories. After a week of the original analysis, each rater reanalysed 50% of the studies (within-coder concordance). The between- and within-rater concordance were estimated with true kappa \bar{k} ; Fariña et al., 2002). This was to correct the Cohen's kappa controlling a systematic source of error: the correspondence between coding (true kappa). Succinctly, if the exact correspondence was not verified, the two errors were encoded as an agreement. This correction is called true kappa. The results showed a total concordance ($\bar{k} = 1$). Additionally, the raters were consistent in other studies i.e., in other contexts (Gancedo et al., 2021). Thus, between- and within-raters and inter-contexts true concordance was verified, and the coding accurately reflected the content of the categories i.e., the coding was reliable (Wicker, 1975). In other words, another trained rater would find the same data.

Data analysis

Since the aim of the meta-analysis design was to determine the probability of implanting false memories (observed proportion, p_o), the effect sizes were calculated with Cohen's h , which is equivalent to d (the results

are shown as d), for differences between proportions and a proportion with a constant. The probability of implanting false memories was not reported directly in the primary studies. Thus, it was computed by raters from the data reported ($p_o = \text{number of rich implanted memories of events}/N$). Therefore, the effect size was computed as an h index, contrasting the observed probabilities of implanted memory with a constant, which provides a conservative estimate of the effect (Arce et al., 2020). The constants were taken as follows (Fandiño et al., 2021): a) a trivial probability (≤ 0.05 ; statistical significant criterion: if the effect was ≤ 0.05 , it was insignificant and if it was higher than 0.05, the observed effect; the correspondent d/h scores for p of 0.05 is 0.10); b) a common probability ($= 0.5$; probable, observed in 50% of the population); and c) a normal probability (≥ 0.90 ; normal, observed in 90% or more of the distribution). The magnitude of the increase or decrease of the effect size (observed contingency) was evaluated in terms of the Effect Incremental Index (EII; Arias et al., 2020) i.e., $oe - c / oe$, where oe is the observed effect and c , the constant.

Having obtained the effect sizes, meta-analyses of random effects were carried out, correcting the effect size for sample error (the lack of reliability of the measurement variable—memory implantation—could not be corrected given that the primary studies did not estimate the reliability of the encoding of memory, but concordance) (Schmidt & Hunter, 2015), consisting in weighting each effect size by its sample i.e., the bare-bones procedure (this procedure produces more accurate

estimates of population SDs). As for this, the sample size weighted mean effect size (d_w); sample size weighted observed variance of d -values (S^2_{obs}); the standard deviation of d -values after removing sampling error variance (SD_{obs}); the variance attributed to sampling error variance (S^2_{se}); the variance attributed to sampling error variance (S^2_{res}); the standard deviation of d -values after removing sampling error (SD_{res}); the percent of observed variance accounted by sampling error (%Var); the 95% confidence interval for d -values (95% CI_d); and the 80% credibility interval (80% CV) were computed. For this the meta-analysis software created by Schmidt and Le (2014) was run.

If the sampling error explained the bulk of the variance ($\geq 60\%$; rule for sampling error as the only artifact; Hunter et al., 1982), then the non-explained variance was not systematic, meaning the results were not mediated by moderators. On the other hand, if the variance due to the sampling error explained less than 60% of the results, these were influenced by the effects of moderators. As for the 95% confidence interval for d , if this interval had no zero, the effect was significant. However, with this procedure trivial effects can appear to be significant (Schmidt & Hunter, 2015). Hence, as a safeguard, the effect should not be trivial (insignificant): an effect was trivial if the observed probability was ≤ 0.05 (0.10 in d/h scores); thus, if the confidence interval observed in d included 0.10, the effect was trivial (Mayorga et al., 2020). In relation to the generalization of the effect, if the 80% credibility interval of d had no zero, the results were generalizable to other studies, the lower limit being the minimum result expected for 90% of the studies. Conversely, if the 80% credibility interval of d passed zero, the results of the studies were contradictory, and, consequently, subject to the effects of moderators (Vilarino et al., 2022).

The magnitude of effect size (d) was interpreted qualitatively as small ($d = 0.20$, an effect size larger than 55.6%), moderate ($d = 0.50$, an effect size larger than 63.7%), large ($d = 0.80$, an effect size larger than 71.6%), or more than large ($d > 1.20$, an effect size larger than 80.2%) (Cohen, 2013; Arce et al., 2015). Additionally, the magnitude was interpreted quantitatively as the Probability of Superiority of the Effect Size (PS_{ES} ; Gancedo et al., 2021). This consists in transforming the observed effect size into a percentile. Thus, meta-analytical results may be interpreted in terms of the probability of superiority of the observed effect size (percentile) over all possible effects (percentile distribution). Finally, the mean effect sizes of the meta-analyses were compared converting effects to correlations and computing the difference between two correlations (q). Then, the Zeta value for the observed difference was estimated with the associated probability (α) (Arias et al., 2020). As for the computation of the PS_{ES} and the comparison of the effects, the authors created excel spreadsheets, the correctness of which were contrasted with a manual execution.

Results

Global analysis

The results of the analysis of the hypothesis of the existence of false memories (in contrast to the probability of observing nothing, 0, i.e., the non-existence of false memories; see Table 2), showed a significant (the confidence interval for d had no 0); positive (supporting false memory implantation); and more than large ($d > 1.20$; an effect size above 84.4%, $PS_{ES} = 0.844$) mean effect size ($d_w = 1.43$). Moreover, it was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.13 (lower limit of the credibility interval), and not subject to moderators (sampling error explained more than 60% of the variance i.e., the studies were homogeneous).

The results of the meta-analysis on the observed probability contrasted with a constant (trivial prevalence; 0.10 in d scores; see Table 2), showed the registered prevalence of false memories was significant, positive (higher than a trivial prevalence), and with a large effect size ($d > 0.80$; an effect size above 75.5%, $PS_{ES} = 0.755$). Moreover, the increase in the effect (false memory implantation) over a trivial effect was 89.8% ($EII = 0.898$). This result was generalizable to 90% of the studies with a minimum effect of 0.67 (lower limit of the credibility interval), and not subject to moderators (%Var ≈ 60).

In the comparison of the observed probability of implanting a false memory (success) with the probability of not implanting a false memory (failure), i.e., $p-q$, the results of the meta-analysis (see Table 2) indicated the mean effect size was not significant (the confidence interval for d had 0), that is, an equal probability of success or failure in implanting false memories. In short, the probability of implanting false memories was 50% i.e., a common effect. Nonetheless, these results were not generalizable (the credibility interval had 0) to inter-studies, and were subject to moderators (%Var < 60). That is, effects under 50% (the lower limit of the credibility interval was -0.48) or above 50% (the upper limit of the credibility interval was 0.34) may be explained by moderators. Bearing in mind that the magnitude of the effects in implanting rich false memories was mediated by moderators, the moderators with implications for the evaluation of the testimony, and the effects were compared.

Study of moderators

Though the previous results corroborated the existence of false memories (i.e., the probability of implanting was significant, generalizable, and not subject to moderators), the literature has focused on the effects of moderators that have an impact on the evaluation of a testimony. A total of four moderators relevant to testimonies have been described in the studies: the type of stimuli (upper limit of the credibility interval of 0.34); the age of the participants (underage and adults); the suggestive techniques (suggestive technique used for the interviewer to introduce

Table 2
Meta-analyses of the implantation of false memories.

k	N	d_w	S^2_{obs}	SD_{obs}	S^2_{se}	%Var	S^2_{res}	SD_{res}	95% CI_d	80% CV
For non-effect (0.0)										
35	1901	1.43	0.1455	0.3815	0.0935	64.27	0.0520	0.2280	1.33, 1.53	1.13, 1.73
For Trivial Effect (0.05)										
35	1901	0.98	0.1439	0.3794	0.0836	58.09	0.0603	0.2456	0.88, 1.08	0.67, 1.29
For a common effect (.50)										
35	1901	-0.07	0.1763	0.4199	0.0746	42.32	0.1017	0.3189	-0.16, 0.02	-0.48, 0.34

Note. k : Number of effect sizes; N : total sample size; d_w : Sample size weighted mean effect size; S^2_{obs} : Sample size weighted observed variance of d -values; SD_{obs} : Sample size weighted observed standard deviation of d -values; S^2_{se} : Variance attributed to sampling error variance; %Var = Percent of observed variance accounted by sampling error variance; S^2_{res} : Variance of d -values after removing sampling error variance; SD_{res} : Standard deviation of d -values after removing sampling error; 95% CI_d : 95% confidence interval for d ; 80% CI_δ = 80% credibility interval.

Table 3
Meta-analyses of the implantation of false memories for the type of stimuli moderator.

<i>k</i>	<i>N</i>	<i>d_w</i>	<i>S²_{obs}</i>	<i>SD_{obs}</i>	<i>S²_{se}</i>	%Var	<i>S²_{res}</i>	<i>SD_{res}</i>	95% CI _{<i>d</i>}	80% CV
Experienced event										
2	148	2.03	0.0331	0.1820	0.0827	100 ¹	0.0000	0.0000	1.63, 2.43	2.03, 2.03
False narrative										
28 ^a	1399	1.35	0.1563	0.3953	0.0996	63.71	0.0567	0.2382	1.23, 1.47	1.04, 1.65
24 ^b	1208	1.33	0.1585	0.3981	0.0984	62.05	0.0601	0.2453	1.21, 1.45	1.02, 1.64
4 ^c	191	1.46	0.1269	0.3562	0.1075	84.71	0.0194	0.1393	1.14, 1.78	1.28, 1.64
Doctored photograph										
6	354	1.29	0.0382	0.1955	0.0831	100 ¹	0.0000	0.0000	1.06, 1.52	1.29, 1.29

Note.

¹ The predicted variance overcomes the observed variance, rounding it to 100%.

^a Narrated or read false narrative.

^b Narrated false narrative.

^c Read false narrative.

the false memory) employed for implanting false memories; and the emotional charge of the event (positive and negative events). As the magnitude of the effect was mediated by moderators (for the contrast of p_o-q , the explained variance by sampling error was less than 60%), the effects on the moderators relevant for forensic practice were analysed and the effects compared.

The primary studies assessed several stimulus types for implanting false memories: experienced event (self-experienced event, excluding events watched on video), false narrative (read by participants or narrated by researchers), and doctored photographs. Thus, a meta-analysis was performed on the effects of each of these moderators (stimulus type). The results of the meta-analysis for ‘experienced event’ stimulus type moderator (see Table 3) found a positive (self-experienced event implant false memories); significant (the confidence interval for *d* had no 0); and more than large ($d > 1.20$; an effect size above 92.5%, $PS_{ES} = 0.925$) mean effect size ($d_w = 2.03$). The generalization to other studies was not possible as the residual variance was zero (i.e., the primary studies were not randomly distributed; $N < 400$ and $k < 3$). Thus, the probability of implanting false memories with this method was significant, but further studies are required to ascertain their full impact.

In relation to the ‘false narrative’ stimulus type, in the primary studies the false event was either read by the participant, or narrated by the researcher. Hence, separated meta-analysis were performed for a false narrative read by the participant, or narrated by the researcher. The results (see Table 3) exhibited equal effect in read by participant ($d_w = 1.46$), or narrated by researcher ($d_w = 1.33$), $q(N = 328) = 0.054, Z = 0.69, p = 0.490$. As the medium for introducing the false event (read vs. narrated) had no different effects in implanting false memories, a meta-analysis was conducted with all the effect sizes of the false event stimulus. The results of the meta-analysis for the moderator ‘false narrative’ stimulus type (see Table 3) revealed a positive (false narrative implanted false memories); significant (the confidence interval for *d* had no 0); and more than large ($d > 1.20$; an effect size above 82.9%, $PS_{ES} = 0.829$) mean effect size ($d_w = 1.35$). This result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.04 (lower limit of the credibility interval); and not subject to moderators (%Var > 60). Thus, the result

was conclusive: the probability of implanting false memories with this method was significant, of an extraordinarily large magnitude, generalizable, and not mediated by moderators, that is, the effect was exclusively due to the method.

The results of the meta-analysis for the moderator ‘doctored photograph’ stimulus type (see Table 3) revealed a positive (doctored photographs implant false memories); significant (the confidence interval for *d* had no 0); and more than large ($d > 1.20$; an effect size above 81.9%, $PS_{ES} = 0.819$) mean effect size ($d_w = 1.29$). The generalization to other studies was not possible as the residual variance was zero (i.e., the primary studies were not randomly distributed; $N < 400$), and additional studies to estimate it are required. In short, the probability of implanting false memories with this method was significant, but more studies are needed to confirm the effects of this moderator.

Comparatively, the ‘self-experienced event’ type of stimuli obtained a significantly higher probability of implanting false memories than the ‘false narrative’, $q(N = 265) = 0.258, Z = 2.95, p = .003$, or ‘doctored photograph’, $q(N = 208) = 0.284, Z = 2.88, p = .004$, stimulus type; whilst ‘false narrative’ and ‘doctored photograph’ stimulus type had the same effect in implanting false memories, $q(N = 563) = 0.029, Z = 0.44, p = .660$.

The results of the meta-analysis for implanting false memories for the age of the participant moderator (see Table 4) showed for the underage a significant (the confidence interval for *d* had no 0); positive (it is probable to implant false memories in children and adolescents); and more than large ($d > 1.20$; an effect size above 84.6%, $PS_{ES} = 0.846$) mean effect size ($d_w = 1.44$). This result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.03 (lower limit of the interval), but subject to moderators (%Var < 60). Thus, the probability of implanting false memories in children and adolescents was significant, generalizable, of an extraordinarily large magnitude, but subject to the effects of moderators. As the primary studies divided their samples into two cohorts (children under 13 years, and adolescents from 13 to 18 years), educational level has a weak effect on rich memories (Álvarez et al., 2021), and that children testimony was highly influenced by suggestibility (Köhnken, 2004), that is, a higher probability of implanted memories, a meta-analysis of children’s

Table 4
Meta-analyses of the development of false memories for the age of the participant moderator.

<i>k</i>	<i>N</i>	<i>d_w</i>	<i>S²_{obs}</i>	<i>SD_{obs}</i>	<i>S²_{se}</i>	%Var	<i>S²_{res}</i>	<i>SD_{res}</i>	95% CI _{<i>d</i>}	80% CV
Underage (< 18 years)										
14	841	1.44	0.1881	0.4337	0.0850	45.18	0.1031	0.3211	1.29, 1.59	1.03, 1.85
13 ^a	790	1.47	0.1893	0.4351	0.0846	44.72	0.1047	0.3235	1.31, 1.63	1.05, 1.89
Adults (≥ 18 years)										
20	1040	1.36	0.1257	0.3545	0.0959	76.28	0.0298	0.1726	1.22, 1.50	1.14, 1.58

Note.

^a Meta-analysis for children (< 13 years).

samples (< 13 years) was replicated, and found similar results (the confidence intervals for *d* overlapped; see Table 4).

The results of the meta-analysis for implanting false memories for the age of the participant moderator (see Table 4) in adults exhibited a significant (the confidence interval for *d* had no 0); positive (it was probable to implant false memories in adults); and more than large (*d* > 1.20; an effect size above 83.2%, $PS_{ES} = 0.832$) mean effect size ($d_w = 1.36$). This result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.14 (lower limit of the credibility interval), and not subject to moderators (%Var > 60). Thus, the result was conclusive: the probability of implanting false memories in adults was significant, of an extraordinarily large magnitude, generalizable, and not mediated by moderators. Comparatively, the probability of implanting false memories in adults and the underage was similar in both, $q(N = 930) = 0.033, Z = 0.71, p = .478$.

The instructions for implanting false memories in the primary studies consisted of 4 techniques: guided imagery (participants guided to imagine events as if they really had happened); prompts/cues (prompts/cues provided by the interviewer to remember a non-recalled event); pressure to answer (participants are urged by the interviewer to recall because their friends, parents or others remembered or already told the event); and a non-directive technique (participants are asked by the interviewer to think about the event).

As in the primary studies the guided imagery suggestive technique for implanting memory was used on its own, or complemented with context reinstatement (guided imagery and context reinstatement), separate meta-analyses were performed to determine if the addition of the context reinstatement instruction (*k* = 6) to the guided imagery technique (*k* = 13) increased the implanting of false memories. The results revealed a similar effect in both conditions (the confidence intervals for *d* overlapped; see Table 5); that is, the results for the guided imagery technique on its own were similar to those obtained for the guided imagery technique complemented with context reinstatement, indicating context reinstatement did not increase the effect of guided imagery. Hence, all of the studies applying the guided imagery technique were analysed.

The results of the meta-analysis of guided imagery instructions for implanting false memories (see Table 5) showed a significant (the confidence interval for *d* had no 0); positive (implanting false memories with prompts/cues was probable); and more than large (*d* > 1.20; an effect size above 84.9%, $PS_{ES} = 0.849$) mean effect size ($d_w = 1.45$). This result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.19, and not subject to moderators (*d* > 1.20; an effect size above 84.9%, $PS_{ES} = 0.849$). Thus, the result was conclusive: the probability of implanting false memories with guided imagery and guided imagery plus context reinstatement was significant,

of an extraordinarily large magnitude, generalizable, but not mediated by moderators, that is, the effect was exclusively due to the method.

The results of the meta-analysis for prompts/cues instructions for the implantation of false memories (see Table 5) showed a significant (the confidence interval for *d* had no 0); positive (implanting false memories with prompts/cues was probable); and more than large (*d* > 1.20; an effect size above 82.6%, $PS_{ES} = 0.826$) mean effect size ($d_w = 1.33$). Moreover, this result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 0.92 (lower limit of the credibility interval), but subject to moderators (%Var < 60). Thus, the probability of implanting false memories with prompts/cues was significant, generalizable, and with an extraordinarily large magnitude, but subject to the effect of moderators.

The results of the meta-analysis of 'pressure to answer' instructions (e.g., peer pressure) for implanting false memories revealed a significant (the confidence interval for *d* had no 0); positive (implanting false memories via guided imagery and context reinstatement was probable); and more than large (*d* > 1.20; an effect size above 86.4%, $PS_{ES} = 0.864$) mean effect size ($d_w = 1.56$). The generalization to other studies could not be assessed due to the zero residual variance (i.e., the primary studies were not randomly distributed; *N* < 400), and further studies are required to estimate it. Thus, the probability of implanting false memories by pressing to answer was significant, but additional studies are needed to fully understand their impact.

The results of the meta-analysis of the non-directive technique for implanting false memories (see Table 5) showed a significant (the confidence interval for *d* had no 0); positive (implanting false memories via guided imagery and context reinstatement was probable); and large (*d* > 0.80; an effect size above 73.9%, $PS_{ES} = 0.739$) mean effect size ($d_w = 0.90$). The generalization to other studies could not be assessed due to the zero residual variance (i.e., the primary studies were not randomly distributed; *N* < 400), and further studies are required to estimate it. Therefore, the probability of implanting false memories without pressure was significant, but additional studies are needed to fully understand their impact.

Comparatively, non-directive instructions (e.g., think hard about this) for implanting memory entailed less implanting of false memories than guided imagery instructions (guided imagery, and guided imagery plus context reinstatement) for implanting false memory, $q(N = 227) = -0.237, Z = -2.51, p = .012$, or pressure to answer, $q(N = 130) = 0.281, Z = 2.24, p = .025$; whereas false memories with non-directive instructions, or prompts/cues were equally implanted, $q(N = 250) = 0.188, Z = 1.63, p = .059$; with guided imagery and prompts/cues, $q(N = 718) = 0.049, Z = 0.93, p = .352$; with guided imagery and pressure, and with prompts/cues and pressure, $q(N = 211) = 0.064, Z = 0.65, p = .516$.

Table 5
Meta-analyses of the development of false memories for the techniques for implantation of false memories.

<i>k</i>	<i>N</i>	<i>d_w</i>	<i>S²_{obs}</i>	<i>SD_{obs}</i>	<i>S²_{se}</i>	%Var	<i>S²_{res}</i>	<i>SD_{res}</i>	95% CI _{<i>d</i>}	80% CV
Guided imagery										
6 ^a	354	1.51	0.0884	0.2974	0.0880	99.55	0.0004	0.0198	1.27, 1.75	1.49, 1.54
13 ^b	754	1.42	0.1473	0.3839	0.0874	59.30	0.0600	0.2449	1.26, 1.58	1.11, 1.73
19 ^c	1108	1.45	0.1303	0.3610	0.0858	67.22	0.0427	0.2067	1.32, 1.58	1.19, 1.71
Prompts/Cues										
8	532	1.33	0.1788	0.4229	0.0743	41.55	0.1045	0.3233	1.14, 1.52	0.92, 1.74
Pressure to answer										
5	133	1.56	0.1853	0.4305	0.2007	100 ¹	0.0000	0.0000	1.17, 1.95	1.56, 1.56
A non-directive technique										
4	128	0.90	0.0245	0.1566	0.1402	100 ¹	0.0000	0.0000	0.53, 1.27	0.90, 0.90

Note.

^a Guided imagery.

^b Guided imagery and context reinstatement.

^c Guided imagery, and guided imagery plus context reinstatement.

¹ The predicted variance overcomes the observed variance, rounding it to 100%.

Table 6
Meta-analyses of the development of false memories for the event emotional valence moderator.

<i>k</i>	<i>N</i>	d_w	S^2_{obs}	SD_{obs}	S^2_{se}	%Var	S^2_{res}	SD_{res}	95% CI_d	80% CV
Positive event										
11	593	1.27	0.1085	0.3295	0.0905	83.39	0.0180	0.1343	1.09, 1.45	1.10, 1.44
Negative event										
22	1126	1.30	0.1384	0.3720	0.0958	69.20	0.0426	0.2065	1.17, 1.43	1.04, 1.56

The results of the meta-analysis of positive events (see Table 6) found a significant (the confidence interval for d had no 0); positive (it is probable to implant false memories of positive emotional events); and more than large ($d > 1.20$; an effect size above 81.6%, $PS_{ES} = 0.816$) mean effect size ($d_w = 1.27$). Moreover, this result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.10 (lower limit of the credibility interval), and was not subject to moderators (%Var > 60). In short, the result was conclusive: the probability of implanting false memories in positive events was significant, of an extraordinarily large magnitude, generalizable, and not mediated by moderators.

As for the meta-analysis of negative events (see Table 6), the results revealed a significant (the confidence interval for d had no 0); positive (implanting false memories of negative emotional events was probable); and more than large ($d > 1.20$; an effect size above 82.1%, $PS_{ES} = 0.821$) mean effect size ($d_w = 1.30$). Moreover, this result was generalizable (the credibility interval had no 0) to 90% of the studies with a minimum effect of 1.04 (lower limit of the credibility interval), and was not subject to moderators (%Var > 60). In short, the result was conclusive: the probability of implanting false memories of negative events was significant, of a magnitude more than large, generalizable, and not mediated by moderators.

The comparison of the results of positive and negative events showed the effects were similar in both, $q(N' = 776) = 0.013$, $Z = 0.26$, $p = .795$.

Discussion

The findings corroborated the hypothesis of the existence of false memories, and the general inter-studies result was of an extraordinarily large effect size (above $\pm 85\%$). Moreover, a File Drawer Analysis (FDA) showed that the findings were resistant to sampling bias ($424 > 185$; Rosenthal, 1979), and to reduce the observed effect of implanting false memories to a trivial effect (insignificant), 0.10 (Fandiño et al., 2021), a total of 465 studies with a null effect would be required to refute the results (Schmidt & Hunter, 2015). Succinctly, strong empirical support (more than large effect size of the probability of implanting false memories); the generalization of the inter-studies results (the studies were homogeneous in implanting false memories); and the absence of studies with inconvenient results (without effect), confer the implanting of false memories the status of 'scientific fact' or 'scientific truth' (Popper, 1973). This result underscored that forensics techniques for evaluating testimonies based on the content analysis of memories should not focus on the classification of honest memories, or in discriminating between honest and dishonest testimonies (false memories are honest testimonies), but rather they should aim to discriminate between memories of self-experienced events, and fabricated memories of events (false memories are fabricated memories, not lived ones). If the forensics tools applied have not been contrasted in the correct classification of implanted memories or in discriminating between memories of self-experienced events or implanted memories of events, they are invalid (e.g., Criteria Based Content Analysis [CBCA]; Steller & Köhnken, 1989; Reality Monitoring [RM]; Gancedo et al., 2021; Sporer & Küpper, 2004). Moreover, in accordance with the principle of the presumption of innocence (Art. 11 of the Universal Declaration of Human Rights; United Nations, 1948), and that no innocent person should be found guilty, the probability of false positives (i.e., implanting memories classified as

self-experienced) should be guaranteed to be 0. However, the probability of implanting false memories was common ($= 0.50$).

As for the stimulus material for implanting memories, these can include an array of stimuli (experienced events, false events, doctored photographs), which is in line with the general results. Though all of the stimulus conditions examined had an impact on judicial testimonies, the implanting of false memories in experienced events entails additional implications for the forensic evaluation of a testimony. Thus, by implanting memories into experienced events, the part of the witness' memory of self-experienced events is truthful. Consequently, the reference forensic techniques for assessing witness credibility focused on classifying truthful statements are based on reality criteria (e.g., CBCA), memory attributes (e.g., RM), or content categories of self-experienced events (e.g., Global Evaluation System [GES]; Arce & Fariña, 2014) to identify this memory as real (CBCA), of external origin (RM), or with content criteria of self-experienced events (GES). This underscores that categorical content analysis should be complemented with an analysis of validity to control this effect. Reality Monitoring only relates memory attributes without proposing a specific forensic technique (e.g., interview, validity check; Gancedo et al., 2021), so this memory of self-experienced events should be evaluated as part of the testimony. The SVA includes the testing of the hypothesis of implanting a false event into a real one, but in spite of the cautioning, there is no means of controlling false memory in the analysis of the testimony (Köhnken, 2004). Regarding GES, content analysis is exclusively restricted to the events being investigated, eliminating from the content analysis all memory that is not the target of the events under investigation. For example, an accusation of sexual assault must be restricted to the content analysis of the assault itself, eliminating prior or post testimony of the assault. In short, the control of the effects of the incrustation of a false memory in an experienced event rest on the analysis of the validity of the evidence.

In the forensic evaluation of testimonies, false memories have been particularly associated to allegations of child sexual abuse (Davis & Loftus, 2020). The results corroborated that the probability of implanting any type of event (implanted events are diverse, the effects of the content of the event have not been analysed, and no study on sexual abuse was found) was similar in both children and adults.

As for the techniques for implanting false memories and their implications in the evaluation of a testimony, the results revealed that the guided imagery technique used by police for recovering memory from witnesses and suspects, and clinicians in recovering traumatic memories (Davis & Loftus, 2020) were associated to a high prevalence of implanted false memories. Likewise, the prompts and cues that are frequently employed in police interviews and in judicial cross-examination also foster the implanting of false memories. Thus, the use of prompts and cues in police interviews should be controlled, and the memory obtained from them deemed invalid. The pressure to answer technique also facilitated the implantation of false memories. The pressure to answer in judicial proceedings comes mainly from police interrogation, judicial cross-examination, and judges' interviews to witness. Hence, a witness' observance of pressure to answer in police interrogation, judicial cross-examination, and judges' interviews invalidates the evidence obtained. Even the use of non-directive techniques employed by judges in judicial cross-examination (Arce, 2017) facilitated the implanting of false memories, but to a lesser extent. Therefore, these memories should be removed as legal evidence from the judicial decision-making process as they constitute extra-legal evidence. Likewise, forensic interview

techniques should not employ prompts or cues (interrogative phase) as they facilitate false memory implantation i.e., contaminates memory (e.g., NICHHD, Poole Guide, stepwise interview, improved cognitive interview). Furthermore, the information obtained in the interviews lack forensic value, as the objective of the evaluation is not to obtain specific information or the cross-examination of the witness (police/judicial aim), but the analysis of content categories of memory. Moreover the free recall forensic interview technique should not be based on guiding imagination i.e., interviewers should not guide witnesses to imagine the event, as a source of contamination of the memory, but in context reinstatement (Fisher & Gieselman, 1992; Memon et al., 2010) of sensory memories of the event (Arce, 2017).

Though false memories have often been related to negative events, the results revealed positive events can be equally implanted. For forensic evaluation, this implies that a memory of a positive event may be implanted, so a false memory can support innocence (e.g., an accused of sexual abuse may implant in child memory a positive event). Thus, in a claimant's non-guilty testimony, the implanting of a positive event should be suspected.

The results of the meta-analyses were subject to several limitations that should be borne in mind. First, though the results substantiated the existence of false memories, the effect sizes observed were so extraordinarily high that it was highly probable they had been overestimated. It is well known that simulation designs (all of the primary studies were based on simulation designs), with judicial implications obtained significantly different results between field and experimental studies (Fariña et al., 1994). In fact, simulation design studies in legal psychology have face validity, but external validity is uncertain (Konecni & Ebbesen, 1992). Second, primary studies did not control the effect of both the interview and the interviewer in memory content. The literature on the effects of the type of interview on memory content is abundant (Memon et al., 2010). Furthermore, interviewers are often untrained and have no access to standardised training schemes. Therefore, two sources of error go uncontrolled: the contamination of memory by the interviewer, and the effect of the interviewer in memory productivity (Arce, 2017). Third, the encoding of memories has been neither systematised, nor responded to categories of content analysis constituting a methodic system (i.e., reliable and valid). Fourth, the agreement between two raters was evaluated, not the reliability of the encodings. Agreement is not reliability. Moreover, as the results were limited to agreement between two raters and non-systematic measures, it was not possible to ascertain if encoding results were generalizable to other raters (i.e., valid). Fifth, in several meta-analyses the primary studies were not randomly distributed (the variance after removing sampling error variance was 0), so the results were estimates. This highlights the need for further studies to substantiate the results of the meta-analyses of the pressure to answer and non-directive techniques to implant a memory; and of the experienced event and doctored photograph stimulus types. Sixth, the designs did not control if the false memory implantation was a real (conversion), or a fake implantation (normative pressure) (Moscovici, 1980). Once again, this overestimated the effect.

Hence, further research should examine the real probability (in field studies) of implanting false memories and the moderators of the effects in prompts/cues instructions and in the underage. Likewise, future literature should seek to overcome the limitations of method concerning the content analysis system, the type of interview, the interviewer, and the reliability and validity of the measures.

Funding

This research has been partially sponsored by a grant of the Ministry of Science and Innovation of Spain (PID2020-115881RB-I00) and by a grant of the Consellería de Cultura, Educación e Ordenación Universitaria of the Xunta de Galicia (ED431B 2020/46).

References marked with an asterisk indicate studies included in the meta-analysis

Declaration of competing interest

The authors declare that there is no conflict of interest.

References

- Álvarez, M. A., Cruz, T., Balmaseda, R., Gracia-Morilla, C., García-Morales, L., Fernández-Cahill, M., & Manzanero, A. L. (2021). Memory of visual paired associates in healthy aging: A transcultural study. *Revista Iberoamericana de Psicología y Salud*, 12(2), 113–120. doi:10.23923/j.rips.2021.02.049.
- 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. doi:10.5944/ap.14.2.21347.
- Arce, R., Arias, E., Novo, M., & Fariña, F. (2020). Are interventions with batterers effective? A meta-analytical review. *Psychosocial Intervention*, 29(3), 153–164. doi:10.5093/pi2020a11.
- Arce, R., & Fariña, F. (2014). Práctica de la prueba psicológica—forense: El Sistema de Evaluación Global (SEG) [Practice of forensic psychological testing: The Global Evaluation System (GES)]. In C. Chan, C. Estrada, F. J. Rodríguez (Eds.), *Aportaciones a la psicología jurídica y forense desde iberoamérica* (pp. 47–61). Manual Moderno.
- Arce, R., Fariña, F., Seijo, D., & Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment*, 22(6), 769–777. doi:10.1177/1073191114558111.
- Arias, E., Arce, R., Vázquez, M. J., & Marcos, V. (2020). Treatment efficacy on the cognitive competence of convicted intimate partner violence offenders. *Anales de Psicología/Annals of Psychology*, 36(3), 427–435. doi:10.6018/analesps.428771.
- *Bruck, M., Ceci, S. J., & Hembrooke, H. (2002). The nature of children's true and false narratives. *Developmental Review*, 22(3), 520–554. doi:10.1016/S0273-2297(02)00006-0.
- *Calado, B., Luke, T. J., Connolly, D. A., Landström, S., & Otgaar, H. (2021). Implanting false autobiographical memories for repeated events. *Memory (Hove, England)*, 29(10), 1320–1341. doi:10.1080/09658211.2021.1981944.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Academic Press.
- Davis, D., & Loftus, E. (2015). Repressed memories. In R. L. Cautin, & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology*. John Wiley and Sons. doi:10.1002/9781118625392.wbecp270.
- Davis, D., & Loftus, E. (2020). Recovered memories and false memories. In J. R. Geddes, N. C. Andreasen, G. M. Goodwin (Eds.), *New oxford textbook of psychiatry* (3rd ed. (pp. 884–893). Oxford University Press. doi:10.1093/med/9780198713005.003.0085.
- Deeb, H., Vrij, A., Leal, S., Fallon, M., Mann, S., Luther, K., & Granhag, P. A. (2022). Mapping details to elicit information and cues to deceit: The effects of map richness. *European Journal of Psychology Applied to Legal Context*, 14(1), 11–19. doi:10.5093/ejpalc2022a2.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17–22. doi:10.1037/h0046671.
- Desjardins, T., & Scoboria, A. (2007). You and your best friend Suzy put slime in Ms. Smollett's desk": Producing false memories with self-relevant details. *Psychonomic Bulletin & Review*, 14(6), 1090–1095. doi:10.3758/bf03193096.
- Fandiño, R., Basanta, J., Sanmarco, J., Arce, R., & Fariña, F. (2021). Evaluation of the executive functioning and psychological adjustment of child to parent offenders: Epidemiology and quantification of harm. *Frontiers in Psychology*, 12, 616855. doi:10.3389/fpsyg.2021.616855.
- Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Lineups: A comparison of high fidelity research and research in a real context]. *Psicothema*, 6(3), 395–402. <http://www.psicothema.com/pdf/935.pdf>.
- Fariña, F., Arce, R., & Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema*, 14(1), 39–46. Retrieved from <http://www.psicothema.com/pdf/684.pdf>.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas.
- Gancedo, Y., Fariña, F., Seijo, D., Vilarinho, M., & Arce, R. (2021). Reality monitoring: A meta-analytical review for forensic practice. *European Journal of Psychology Applied to Legal Context*, 13(2), 99–110. doi:10.5093/ejpalc2021a10.
- *Garry, M., & Wade, K. (2005). Actually, a picture is worth less than 45 words: Narratives produce more false memories than photographs do. *Psychonomic Bulletin & Review*, 12(2), 359–366. doi:10.3758/BF03196385.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. John Wiley & Sons.
- *Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 920–930. doi:10.1037/0278-7393.27.4.920.
- *Hessen-Kayfitz, J. K., & Scoboria, A. (2011). False memory is in the details: Photographic details differentially predict memory formation. *Applied Cognitive Psychology*, 26(3), 333–341. doi:10.1002/acp.1839.
- *Huffman, M. L., Crossman, A. M., & Ceci, S. L. (1997). "Are false memories permanent?": An investigation of the long-term effects of source misattributions. *Consciousness and Cognition*, 6(4), 482–490. doi:10.1006/ccog.1997.0316.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Sage.

- Hyman, I. E., & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory (Hove, England)*, 6(1), 1–20. doi:10.1080/741941598.
- *Hyman, I. E., Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology*, 9(3), 181–197. doi:10.1002/acp.2350090302.
- *Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language*, 35(2), 101–117. doi:10.1006/jmla.1996.0006.
- Köhnken, S. (2004). Statement reality analysis and the detection of the truth. In P. A. Granhag, & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge University Press. doi:10.1017/CBO9780511490071.003.
- Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime & Law*, 5(1–2), 3–27. doi:10.1080/10683169908414991.
- Konecni, V. J., & Ebbesen, E. B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, T. Bliessner (Eds.), *Psychology and law: International perspectives* (pp. 413–423). Walter de Gruyter.
- Lindsay, D. S., & Read, J. D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology*, 8(4), 281–338. doi:10.1002/acp.2350080403.
- *Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science*, 15(3), 149–154. doi:10.1111/j.0956-7976.2004.01503002.x.
- Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory*. St. Martin's Press.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4(1), 19–31. doi:10.1037/0278-7393.4.1.19.
- *Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. doi:10.3928/0048-5713-19951201-07.
- Mayorga, E. G., Novo, M., Fariña, F., & Arce, R. (2020). Destrezas cognitivas en menores infractores, de protección y normalizados: Un estudio de contraste [Cognitive skills in juvenile offenders, protection and normalized youngsters: A contrastive study]. *Revista Latinoamericana de Psicología*, 52, 160–168. doi:10.14349/rfp.2020.v52.16.
- *Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science*, 14(2), 186–188. doi:10.1046/j.1432-1327.1999.00020.x.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). Cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16(4), 340–372. doi:10.1037/a0020518.
- Memon, A., Wark, L., Bull, R., & Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, 88(2), 179–198. doi:10.1111/j.2044-8295.1997.tb02629.x.
- Moscovici, S. (1980). Toward a theory of conversion behaviour. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 209–239). Academic Press (vol. 13).
- Novo, M., & Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *European Journal of Psychology Applied to Legal Context*, 2(2), 91–115. http://sepjf.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=26&Itemid=110&lang=es.
- Ost, J., Foster, S., Costall, A., & Bull, R. (2005). False reports of childhood events in appropriate interviews. *Memory (Hove, England)*, 13(7), 700–710. doi:10.1080/09658210444000340.
- Otgaar, H., & Candel, I. (2011). Children's false memories: Different false memory paradigms reveal different results. *Psychology, Crime & Law*, 17(6), 513–528. doi:10.1080/10683160903373392.
- *Otgaar, H., Candel, I., & Merckelbach, H. (2008). Children's false memories: Easier to elicit for a negative than for a neutral event. *Acta Psychologica*, 128(2), 350–354. doi:10.1016/j.actpsy.2008.03.009.
- *Otgaar, H., Candel, I., Merckelbach, H., & Wade, K. (2009). Abducted by a UFO: Prevalence information affects young children's false memories for an implausible event. *Applied Cognitive Psychology*, 23(1), 115–125. doi:10.1002/acp.1445.
- *Otgaar, H., Candel, I., Scoboria, A., & Merckelbach, H. (2010). Script knowledge enhances the development of children's false memories. *Acta Psychologica*, 133(1), 57–63. doi:10.1016/j.actpsy.2009.09.002.
- *Otgaar, H., Verschuere, B., Meijer, E. H., & van Oorsouw, K. (2012). The origin of children's implanted false memories: Memory traces or compliance? *Acta Psychologica*, 139(3), 397–403. doi:10.1016/j.actpsy.2012.01.002.
- Patihis, L., Ho, L. Y., Tingen, I. W., Lilienfeld, S. O., & Loftus, E. F. (2014). Are the "Memory Wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science*, 25(2), 519–530. doi:10.1177/0956797613510718.
- *Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, 8(6), 437–441. doi:10.1111/j.1467-9280.1997.tb00457.x.
- *Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development*, 70(4), 887–895. doi:10.1111/1467-8624.00064.
- Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study "False memory," and what are the implications of these choices? *Consciousness and Cognition*, 16(1), 2–17. doi:10.1016/j.concog.2005.06.006.
- Popper, K. R. (1973). La lógica de las ciencias sociales [The logic of social sciences]. In T. Adorno, & J. Habermas (Eds.), *La disputa del positivismo en la sociología alemana* (pp. 101–119). Grijalbo.
- *Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, 23(5), 517–537. doi:10.1023/A:1022344128649.
- *Principe, G. F., DiPuppo, J., & Gammel, J. (2013). Effects of mothers' conversation style and receipt of misinformation on children's event reports. *Cognitive Development*, 28(3), 260–271. doi:10.1016/j.cogdev.2013.01.012.
- *Principe, G. F., Trumbull, J., Gardner, G., Van Horn, E., & Dean, A. M. (2017). The role of maternal elaborative structure and control in children's memory and suggestibility for a past event. *Journal of Experimental Child Psychology*, 163(11), 15–31. doi:10.1016/j.jecp.2017.06.001.
- *Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M. E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology*, 72(4), 235–270. doi:10.1006/jecp.1999.2491.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(4), 803–814. doi:10.1037/0278-7393.21.4.803.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. doi:10.1037/0033-2909.86.3.638.
- Schmidt, F. L., & Le, H. (2014). *Software for the Hunter-Schmidt meta-analysis methods. Version 2.0. Department of management and organizations*. University of Iowa.
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting errors and bias in research findings* (3rd ed.). Sage.
- *Shaw, J., & Porter, S. (2018). Corrigendum: Constructing rich false memories of committing crime. *Psychological Science*, 29(4), 673–674. doi:10.1177/0956797614562862.
- Sporer, S. L., & Kupper, B. (2004). Fantasie und Wirklichkeit - erinnerungsqualitäten von erlebten und erfundenen geschichten [Fantasy and reality - memory qualities of true and invented stories]. *Zeitschrift für Psychologie*, 212(3), 135–151. doi:10.1026/0044-3409.212.3.135.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). Springer-Verlag.
- *Strange, D., Hayne, H., & Garry, M. (2008). A photo, a suggestion, a false memory. *Applied Cognitive Psychology*, 22(5), 587–603. doi:10.1002/acp.1390.
- *Strange, D., Sutherland, R., & Garry, M. (2006). Event plausibility does not determine children's false memories. *Memory (Hove, England)*, 14(8), 937–951. doi:10.1080/09658210600896105.
- *Strange, D., Wade, K., & Hayne, H. (2008). Creating false memories for events that occurred before versus after the offset of childhood amnesia. *Memory (Hove, England)*, 16(5), 475–484. doi:10.1080/09658210802059049.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen [Assessing the credibility of witnesses]. In U. Undeutsch (Eds.), *Handbuch der psychologie: Vol (pp. 26–181). II: Forensische psychologie*.
- United Nations. (1948). Universal declaration of human rights. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- Vilariño, M., Amado, B. G., Seijo, D., Selaya, A., & Arce, R. (2022). Consequences of child maltreatment victimization in internalizing and externalizing mental health problems. *Legal and Criminological Psychology*, 27, 182–193. doi:10.1111/lcrp.12212.
- Vilariño, M., Arce, R., & Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *European Journal of Psychology Applied to Legal Context*, 5(1), 1–21. Retrieved from <https://journals.copmadrid.org/ejpalc/archivos/articulo20180219092235.pdf>.
- Vrij, A., Palena, N., Leal, S., & Caso, L. (2021). The relationship between complications, common knowledge details and self-handicapping strategies and veracity: A meta-analysis. *European Journal of Psychology Applied to Legal Context*, 13(2), 55–77. doi:10.5093/ejpalc2021a7.
- *Wade, K. A., Garry, M., Nash, R. A., & Harper, D. N. (2010). Anchoring effects in the development of false childhood memories. *Psychonomic Bulletin & Review*, 17(1), 66–72. doi:10.3758/PBR.17.1.66.
- *Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin & Review*, 9(3), 597–603. doi:10.3758/BF03196318.
- Wicker, A. W. (1975). An application of a multitrait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 1(4), 575–579. doi:10.1177/014616727500100405.
- Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, 7, 294–300. doi:10.1111/j.1467-9280.1996.tb00377.x.