
RAMÓN ARCE

EXTENDED SUMMARY

Introduction

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

Evaluation of the Undeutsch hypothesis under the Daubert standard

In the case of Daubert v. Merrell Dow Pharmaceuticals (1993), the US Supreme Court established the criteria for the admission of expert testimony as scientific evidence in a court of law, which has also been tested in scientific research: 1) Is the scientific hypothesis/technique testable? Not only is it possible to test it, it has already been tested in both experimental (laboratory) studies, and field studies. As for the validation of a theory (hypothesis), the US Supreme Court ruled that validating a theory only on the findings of experimental studies was insufficient, and it should be further substantiated by field studies. Both of the study designs have advantages and shortcomings. The advantage of experimental studies is the high degree of control over the independent variable (causal) manipulated by the researcher, meaning the reliability of the classification of real or fabricated accounts is total at the cost of external validity, particularly in the field of research on the psychology of testimonies. Thus, a study comparing testimonies in a high fidelity experimental study and a field study found inconsistent results (Fariña, Real, & Arce, 1992). The strength of field studies lies in the external validity of the designs, but their weakness is internal validity as they lack control over the manipulated variables i.e., the classification criterion (ground truth) of real events. 2) Has the hypothesis/technique been tested? It has been widely tested and validated by numerous studies (i.e., it is statistically significant in discriminating between memories of real-life events and fabricated accounts), both in experimental studies (δ = 0.56 and 0.32, in populations of children and adults, respectively), and field studies (δ = 2.71 and 0.45; in children and adults, respectively); in plaintiff testimonies in populations of children (δ = 0.79), and adults (δ = 0.56); in eyewitnesses (δ = 0.51) and defendants (δ = 0.35); in an array of crimes (e.g., child sex abuse, δ = 0.79; sexual assault and gender violence, δ = 0.87); and in several contexts (memories of criminal and non-criminal events; memories of fabricated events, induced or suggested accounts) (Amado et al., 2015, 2016). The validity of the hypothesis in both field and experimental studies underscores its robust study design (convergent validity). In other words, it is robust to sources of error.
linked to the design of experimental studies (external validity), and field studies (internal validity). Moreover, it is equally valid in different contexts (i.e., types of crimes), and exhibits inter-context consistency. In addition, the hypothesis has not been rejected by inconvenient results i.e., negative or adverse to the hypothesis mean effect sizes. Likewise, the ground truth factor does not mediate effects in results acquiring consistency among ground truth standards. Thus, the hypothesis has not only been tested and substantiated, it has also been validated and is powerful (moderate, large and more than large effect sizes in different settings). 3) Is there a known error rate? Given that the hypothesis was formulated but not corroborated by the author, that the categorial systems were created on the basis of the authors’ experiences (top-down) derived from thousands of cases without a strict and objective decision criterion, and that it was not tested, would explain why the error rate was not published. Nevertheless, the meta-analysis of Amado et al. (2015, 2016) calculated the statistical margin of error following a systematic analysis of the literature to be around 31.5% of false negatives in the population of children and 36.5% in adults. 4) Has the hypothesis/technique been subjected to peer review and publication (it is not a prerequisite ‘sine qua non’)? On more than 800 occasions (effect size) (Amado et al., 2015, 2016). 5) Is the hypothesis upon which the technique rests on generally accepted as appropriate by the scientific community? The US Supreme Court resorts to the scientific community to attest the method and principles of the hypothesis. To our knowledge, there has been no surveys regarding this or any other psychological hypothesis. From a scientific point of view, the means for consulting the scientific community involves a systematic, meta-analytical review of the literature to determine broad support for the hypothesis i.e., validation by the scientific community (general acceptance). The meta-analysis of Amado et al. (2015, 2016) not only reveals general acceptance, but also found no meta-analytical result (i.e., mean effect size) contradicting the hypothesis.

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

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

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

The design of the categorial system was in accordance with the guidelines of Bardin (1996) to ensure it was methodic i.e., reliable and valid: mutual exclusion, homogeneity, objective, adequate and relevant, exhaustive, fidelity and productivity. The resulting categorial system (see Table 7) was internally consistent (homogeneity), \( \alpha = .789 \) and .856 (Arce, Fariña, Novo, & Vilariño, 2013; Vilariño, Novo, & Seijo, 2011). Moreover, the categories discriminated (discriminant validity) significantly between memories of real-life events and fabricated accounts (Arce et al., 2013; Arce, Fariña, & Vivero, 2007; Vilariño, 2010; Vilariño et al., 2011). A strict decision criterion was empirically (objectively) defined (criterion validity with false positives = 0; principle of the presumption of innocence); a correct execution control system was designed (fidelity) of the measure in the designs of \( N = 1 \) (inter- and intra-rater, and intra-context consistency) and to overcome the deficiencies of SVA. In fact, the Global Evaluation System is a forensic technique that must be carried out by a trained expert; statements must be recorded and archived (external review of the methods used and judicial validity of the evidence); the coding of statements must be registered and archived (external review of the registered criteria and their reliability); codings must be carried out independently by two experts (evaluation of the reliability of the measure); two statements are obtained (evaluation of the consistency of the evidence); and that statements are obtained through a cognitive interview –control of the contamination of evidence resulting from the interrogation--: free-recall and techniques to aid memory recall (without interrogation). The evaluation of the validity of the testimony (validity of the evidence) is ensured if the following 5 criteria based on the analysis of judicial judgements are met (Arce, Seijo, & Novo, 2010): internal consistency (Are there internal contradictions in the account?); external consistency (Is it consistent with robust or other irrefutable evidence?); consistency with prior and subsequent testimonies (Is there consistency between the central aspects and actions of the event?); persistence (Are statements stable in time in the contexts?); and consistent with scientific laws and the laws of nature (Does the account mention events incompatible with scientific laws and the laws of nature?). If the evidence fails to meet any of these validity criteria, no content analysis is undertaken as the evidence is invalid. Moreover, there should be sufficient evidence for it to be the object of content analysis, sufficient evidence (Is it beyond the witness’ memory capacity? Does it contain all the necessary information regarding the events?).

Further lines of research

Taking into account content categories discriminate between memories of self-experienced real-life events versus fabricated or fictitious accounts; that a methodic categorial systems may be designed; that there can be more categories, mainly those specific to contexts of victimization i.e., school bullying, gender violence, harassment at work, sexual assault in adults (exhaustive), research should seek to detect and analyse new content categories for the previously mentioned contexts. The most efficacious and productive method entails successive approximations (Fariña, Arce, & Novo, 2002) based on statements from judicial cases and fabricated
accounts, and applying the techniques described to the categories identified in accordance with the guidelines of Bardin for designing a methodic categoria system. Furthermore, a strict decision criterion (to ensure the number of false positives is 0) must be continuously revised since lying is also subject to continuous evolution (thus the importance of not revealing intricacies of the technique to prevent the learning of the technique); in comparison, honesty always remains constant.